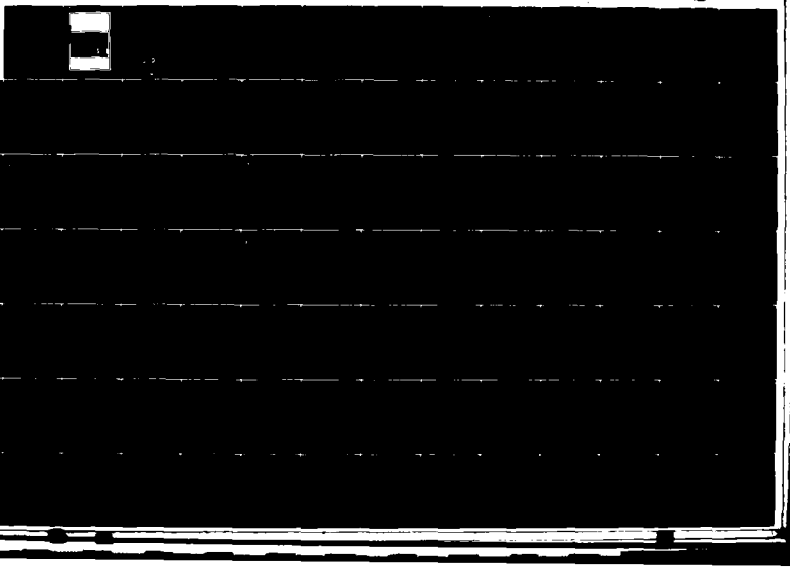
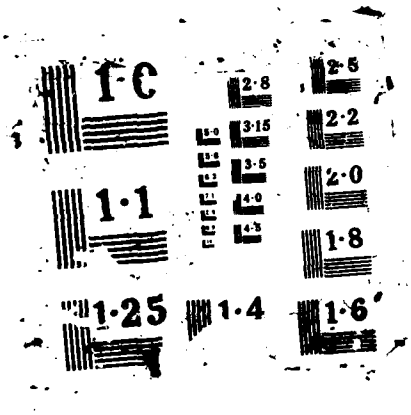
AD-A194 653 BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION CO..(U) ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT NEUILLY.. MAR 88 AGARD-CP-438 1/2

UNCLASSIFIED F/G 5/2 NL




1·C
1·1
1·25
2·8
3·15
3·5
4·0
4·5
1·4
2·5
2·2
2·0
1·8
1·6

AGARD-CP-430

AGARD

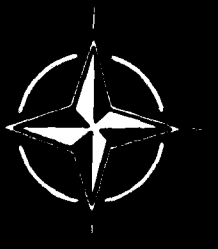
ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No.430

Barriers to Information Transfer and Approaches Toward their Reduction

NORTH ATLANTIC TREATY ORGANIZATION



DISTRIBUTION AND AVAILABILITY ON BACK COVER

COMPONENT PART NOTICE

THIS PAPER IS A COMPONENT PART OF THE FOLLOWING COMPILATION REPORT:

TITLE: Barriers to Information Transfer and Approaches Toward Their Reduction: Conference Proceedings of the Technical Information Panel Specialists' Meeting Held in Washington, DC on 23-24 September 1987.

TO ORDER THE COMPLETE COMPILATION REPORT, USE AD-A194 653.

THE COMPONENT PART IS PROVIDED HERE TO ALLOW USERS ACCESS TO INDIVIDUALLY AUTHORED SECTIONS OF PROCEEDING, ANNALS, SYMPOSIA, ETC. HOWEVER, THE COMPONENT SHOULD BE CONSIDERED WITHIN THE CONTEXT OF THE OVERALL COMPILATION REPORT AND NOT AS A STAND-ALONE TECHNICAL REPORT.

THE FOLLOWING COMPONENT PART NUMBERS COMPRISE THE COMPILATION REPORT:

AD#: AD-P005 721 thru AD#: AD-P005 730

AD#: AD#:

AD#: AD#:

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC ELECTE JUL 07 1988 S D E

DTIC FORM 463 MAR 85

OPI: DTIC-TID

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.430

BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Washington, DC, USA on 23—24 September 1987.

THE MISSION OF AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;
- Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Exchange of scientific and technical information;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

Published March 1988

ISBN 92-835-0449-6

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

THEME

Barriers to information transfer are a significant factor in the aerospace and defence R & D community. This community most often uses the scientific and technical information of its own nation, although the literature originates worldwide. Understanding these barriers will help identify the best applications of information technology for barrier reduction.

The programme sought first to define the magnitude of the problem in terms of aerospace and defence information-seeking behaviour, and to identify the language, culture and economic barriers to information transfer.

The next sessions explored the man-machine issues, such as the viability of computer-assisted translation, and machine and data issues, such as the availability and need for standards. The status and potential application of information technology in these areas was also discussed.

* * *

Les barrières au transfert des informations dans la Communauté Aérospatiale de Recherche et de Développement posent un réel problème car cette communauté utilise le plus souvent les informations scientifiques et techniques de son pays; bien que la littérature technique soit universelle.

La compréhension de ces barrières permettra de déterminer les meilleures applications aux technologies d'information pour les réduire.

Le programme a cherché, en premier, à définir l'amplitude du problème parmi les chercheurs d'information dans les domaines aérospatiales et de défense et à identifier les barrières économiques, de langage et culturelles au transfert des informations.

Les sessions suivantes ont exploré les problèmes homme-machine telle que la viabilité de l'assistance des computeurs pour traductions; et les problèmes de machines et de données telle que la disponibilité et la nécessité de standards. L'état présent et le potentiel de l'application des technologies pour l'information dans ces domaines ont été aussi discutés.

TECHNICAL INFORMATION PANEL

Chairman: Mr Walter R.Blados
USAF Scientific & Technical Information
Program Manager
Office of the Secretary of the Air Force
SAF/AQXT, Room 4C258, The Pentagon
Washington DC 20330-1000
USA

Deputy Chairman: Lt. Col. Armand Cuffez
Adjoint Informatique Logistique
Etat-Major Force Aérienne (VSL/I)
Quartier Reine Elisabeth
Rue D'Evère
1140 Brussels
Belgium

PROGRAMME AND PLANNING COMMITTEE

Mr Walter R.Blados
USAF Scientific & Technical Information Program Manager
Office of the Secretary of the Air Force
SAF/AQXT, Room 4C258, The Pentagon
Washington DC 20330-1000
USA

Lt. Col. Armand Cuffez
Adjoint Informatique Logistique
Etat-Major Force Aérienne (VSL/I)
Quartier Reine Elisabeth
Rue D'Evère
1140 Brussels
Belgium

Col. Doğan Kaya
Ministry of National Defence
Dept of R&D (ARGE)
Ankara, Turkey

Ms Barbara Lawrence, Administrator
Technical Information Service
American Institute of Aeronautics and Astronautics
555 West 57th Street, 12th Floor
New York, NY 10019
USA

Ir. Axel S.T.Tan
Information Specialist
National Aerospace Laboratory (NLR)
P.O. Box 90502
1006 BM Amsterdam
The Netherlands

Mr Gerd Tittlbach
Fachinformationszentrum
Karlsruhe
D-7514 Eggenstein-Leopoldshafen 2
Federal Republic of Germany

Mr Van A.Wente
Director, Scientific & Technical Information Division
NASA Headquarters (Code NTT)
Washington DC 20546
USA

Miss Nancy M.Wildgoose
Director, Scientific Information Services
Department of National Defence
Ottawa, Ontario K1A 0K2
Canada

MEETING DIRECTOR

Mr Walter R.Blados
USAF Scientific & Technical Information Program Manager
Office of the Secretary of the Air Force
SAF/AQXT, Room 4C258, The Pentagon
Washington DC 20330-1000
USA

THEME COORDINATOR

Ms Barbara Lawrence, Administrator
Technical Information Service
American Institute of Aeronautics and Astronautics
555 West 57th Street, 12th Floor
New York, NY 10019
USA

MEETING LOCAL COORDINATORS

Mr Kurt N.Molholm
Administrator
Defense Technical Information Center
Cameron Station
Alexandria, VA 22304-6145
USA

Mr Van A.Wente
Director, Scientific & Technical
Information Division
NASA Headquarters (Code NTT)
Washington DC 20546
USA

PANEL EXECUTIVE

Mr G.W.Hart
AGARD/NATO
7 rue Ancelle
92200 Neuilly-sur-Seine
France

From USA and Canada:
AGARD/NATO
APO New York 09777
Attention: TIP

CONTENTS

*Papers 2 and 8 have been translated into English and Paper 4 into French by Aérospatiale, using the machine translation system SYSTRAN with post-editing. Notes on the translations and some examples of the un-edited translations appear at Item T.

BARRIERS TO THE INTERNATIONAL TRANSFER OF INFORMATION IN AEROSPACE AND DEFENSE

by

James J.Harford
Barbara Lawrence
American Institute of Aeronautics and Astronautics
370 L'Enfant Promenade, SW
Washington, D.C. 20024

INTRODUCTION

Let's take our keynote for this meeting from a recent study of the US National Academy of Engineering (NAE), entitled "Strengthening US Engineering Through International Cooperation", (1):

> "The world in which US engineers and technologists learn and practice their profession is changing more rapidly than the institutions, policies, and programs intended to ensure our future economic growth, security, and welfare. The United States and especially its engineering and science community needs to adopt new attitudes and strategies if we are to maintain or enhance our industrial health and standard of living in the face of the reality of intense international economic competition. Responses must be political, economic, and cultural, as well as technological. Although improvements in the competitive status of the United States will not come about solely as a result of our being more aware of technological progress made outside the United States, technological isolation will surely undermine the future of our industrial and educational institutions."

The United States may not be the only NATO nation with isolationist technological tendencies, but it probably has those tendencies to a greater degree than most others, because for so long we have been the undisputed technological leader, while we have been the most backward linguistically.

Professional societies have a special role in correcting this most unsatisfactory situation. Indeed, as organizations chartered to promote information exchange, we have a deep obligation to help import information, and encourage international interactions and awareness, to a much greater degree than in the past. To do this we should first understand the barriers to the global flow of information.

This morning I'll be talking about how international information transfer occurs in aerospace and defense and I will try to characterize the barriers specific to aerospace and defence. Then I will challenge you to participate in the main mission of this meeting — the development of an action program for both AGARD and your own organizations that will reduce the barriers.

ROLE OF A PROFESSIONAL SOCIETY

AIAA has taken to heart the recommendation of the above-mentioned NAE study that "US-based engineering and professional societies should continue to develop their international activities for the benefit of both the US and international engineering and technology enterprise". (2)

We are the oldest and largest professional society of aerospace scientists and engineers in the world. Currently we have 40,000 members. Although only 5% are foreign, the percentage is growing strongly. Our information transfer activities fall into three areas — meetings, publications, and education.

AIAA holds more than 25 technical meetings each year and assists in organizing several international conferences [e.g. the annual congresses of the International Astronautical Federation (IAF), and the biennial International Congress of the Aeronautical Sciences (ICAS)].

Through the good work of members of our Technical Committees efforts to increase international participation in our domestic meetings have been quite successful. Last year's Aerospace Sciences meeting, for example, had 11% non-US papers, the Communications Satellite Systems Conference had 42%, and the Structures, Structural Dynamics and Materials meeting had 10%.

Even the balance of international participation in AIAA meetings has changed. Earlier, English-speaking Western European nations dominated. Now our meetings regularly include papers from the USSR,China, Latin America, and above all Japan. The increase in Japanese participation is dramatic. In the 1980 Aerospace Sciences meeting there were no Japanese papers. In 1987, 20% of the foreign papers were from Japan.

AIAA's Publications Committee has also taken up the call to do its part to improve the diffusion of international information. *Aerospace America*, AIAA's membership magazine, helps readers maintain currency through monthly reporting on aerospace trends in Europe, Japan and the Soviet Union.

One journal, the *Journal of Aircraft*, has an international board of editors whose primary role is to solicit non-US manuscripts. The Publications Committee recently voted to encourage the editors-in-chief of the other AIAA journals to do likewise.

Soviet Abstracts are a new feature in the basic research journal, the *AIAA Journal*. Each month the Journal publishes a minibibliography of abstracts of recent Soviet work on some timely topic.

AIAA's NASA-supported secondary information service publishes the bi-weekly *International Aerospace Abstracts (IAA)* and markets the *Aerospace Database*, a major information resource for international literature in aerospace and related technologies. The sources analysed for these services come from over 100 countries and are in 30 languages. Since one half of the aerospace literature information originates outside the US, it is clear that the need to be aware of worldwide literature is not an idle notion.

AIAA is experimenting with computer-based continuing education, coordinating homework on PCs with electronic mail and teleeducation eliminates the need for travel, and opens the door to international participation — both for students and teachers.

MOTIVES FOR INTERNATIONAL COVERAGE

Knowing the motives for such efforts at international coverage may help us later when we address ways to break down the barriers. Wallerstein, in a National Research Council study entitled "Scientific and Technical Cooperation Among Industrialized Nations" (3), identified several motives for international cooperation:

Cost Sharing — In the current economy, this is perhaps the most pressing argument for international cooperation. Dan Greenberg, writing about funding for the superconducting supercollider says, "since high energy physics is an open, fundamental field of science...., the supercollider should be internationalized and its costs spread among a consortium of scientifically advanced nations." (4) Don Fink, editor of Aviation Week, comes to the same conclusion writing about the recent cancellation of the Israeli fighter plane, "The Lavi cancellation cannot be considered a strictly Israeli problem. It also must be looked upon as the wave of the future for all the Free World nations. The financial realities of the modern world could well mean that international cooperative efforts are the only practical way their military requirements can be met." (5)

Concept Development — formal exploitation of the invisible college

Enhancement of Scientific and Engineering Competence

Political Considerations — enhance alliances, assist developing nations, gain prestige

The other objectives seem to have relevance now:

Defeating the "Not-Invented-Here" (NIH) (6) syndrome — or preferring the arrogant, know it all view

Supporting Development of New Technological Industries — as the EEC has done, when it is clear that, particularly for information technology based products, sticking to national boundaries means markets that are too small.

Let us remember these motives, and consider them when we seek solutions to reducing the international information transfer barriers.

HOW INTERNATIONAL INFORMATION TRANSFER OCCURS

Some avenues of information transfer are specific to the concept international; others are generic.

The aerospace and defense world is primarily a community of engineers. And engineers are notorious for preferring the route of informal communication above all. To an engineer the maxim that "the accessibility of information is more important than the perceived value" (7) is in truth a dictum. Invisible colleges are real institutions, and the telephone stretches their boundaries. Indeed, electronic mail networks, such as ARPANET, function as the infrastructure of the invisible college in its open version.

One way to join the informal network, the invisible college, is by attending meetings and symposia. Various international conferences provide a significant channel of international information exchange, for example the previously mentioned International Astronautical Federation (IAF) and International Congress of the Aeronautical Sciences (ICAS), as well as the Committee on Data for Science and Technology (CODATA).

The final generic information transfer avenue is, of course, the literature. In aerospace and defense the important primary literature includes both published literature and technical report literature. The NASA and AIAA programs emphasize international exchange agreements to ensure building the most comprehensive aerospace information resources. This audience needs no further details, but I will remind you that the very existence of the NATO AGARD Technical Information Panel (TIP) attests to the multinational original of the literature.

Academic exchanges can be used to foster international information flow. Other nations use this route. In the US, in fact, well over 50% of engineering graduate students are foreign. (8)

Sabbaticals and post-doctoral fellowships are the traditional means of academic exchange. However, in the US the

opportunities have decreased significantly. In the early 1960s more than 50% of National Science Foundation (NSF) postdoctoral fellows went to foreign institutions. By the early 1980s the absolute number had dropped to one fifth, and only 10% of the fellows went to foreign institutions.

Academic exchanges rate high on the list of methods of international cooperation among NSF engineering program managers. They also rate high within NATO. Since the beginning of the 1960s the NATO Science Fellowship program has supported 150,000 scientists and engineers in academic exchanges throughout the Alliance. (9)

Thus, international organizations play an important role in facilitating international information transfer. The NATO science program is certainly a leader. NATO has not only the Feliowship program, but also the Research Collaborative Grants program, and the very special Advanced Study Institutes.

AGARD, with its subject expert panel structure provides extended opportunities for formal and informal exchange.

Bilateral programs are still another cross national channel for information transfer. For example, the US has used bilateral arrangements between NASA and ESA, which includes the technical information exchange arrangements that provide reports input to the abstract publication, STAR.

Lastly, industry should not be forgotten in a reveiw of international exchange. Some companies themselves are multinational, with research and development centers around the world; IBM being perhaps the premier example. Companies may license technology from elsewhere. US companies have licensed Soviet welding technology. Also, industry may engage in joint ventures. For instance, Boeing has joined with both Japanese and Italian companies to manufacture the 767 transport aircraft.

TYPES OF BARRIERS

With that as a background, let us review the inhibitors to information exchange — the barriers — in preparation for their detailed discussion during the next two days. The first are cultural barriers, including the aforementioned Not-Invented-Here Syndrome, or technological arrogance. These seem to stem from the societal or organizational mores of some scientists and engineers who simply don't take into account the foreign literature, perhaps because of a lack of education in the use and value of information.

A second major barrier is economic — an alleged defense against competitors. This can, and often does, backfire, as it has for the US semiconductor industry.

A third barrier is language — already referred to and certainly the most obvious challenge to a global community. J.Michel (10) has said that in science and technology there are only five languages of significance — English, Russian, German, French, and Japanese. In aerospace we should now add a sixth language, Chinese. Van Bergeijk (11) of the International Translation Center said, in 1980, that only 25—30% of important literature is translated. We will hear more about the problems and the proposed solutions regarding language barriers in the papers that follow.

Political barriers often result from national pride — we'll do it ourselves — or as in the case of the Israeli Lavi fighter, we can do anything — or from the desire to protect national security. I won't belabor these isuses, but wish to quote from the testimony of Lew Allen in presenting the study "Balancing the National Interest: US National Security Export Controls and Global Economic Competition". (12)

> "... the US must continue to encourage vigorous technology advance in this country, and indeed among our Western allies as well. Economic vitality generally, and economic strength in the high-technology sector particularly, is important for national security."

It sometimes seems that the more we develop and depend on the use of technology to support information transfer, the more barriers we create. Questions of access, training, cost, and standards will also be discussed in depth later in this symposium.

AEROSPACE AND DEFENSE NATURE OF INFORMATION BEHAVIOR

What makes us think that the aerospace community needs to make a special effort at reducing the barriers to international information transfer? There is some evidence that the nature of the large and complex and often sensitive projects in this industry contribute to the narrowing of visible horizons, and the loss of perceived need to look beyond the current program and colleagues at hand.

Building a missile, an airplane, or a space station requires several thousand engineers, all working for several years against predefined goals and specifications. There is a can't-fail attitude that may limit risk taking. In this environment it is not surprising that many of our finest engineers remain bound by the vision of the current project and are not concerned about absorbing new information.

How large is the problem? Let me start with an anecdote. A few years ago the AIAA Board of Directors felt that the need for the US aerospace community to be more conscious of foreign literature had reached an urgent level. Therefore, we began to publish a low-priced topical abstract bulletin of Soviet and Japanese literature. Because sales were slow, and a majority of

orders were not from the aerospace industry, we cancelled the program. We then conducted a telephone survey of AIAA members to udnerstand why interest was so low. The response came back, "Sure, following Soviet and Japanese science and technology is an important issue, but it is not my job". If a senior materials engineer does not think it is his or her job to follow the foreign sci/tech literature, we certainly do have a barrier. (13)

Some comments on the international nature of science and technology today, based on data from the latest National Science Foundation Science Indicators (14) and data we have gathered at AIAA. (15)

— R&D is increasing in all major countries. Although this is true in absolute dollars for all countries, the rate of increase is larger outside the US. In fact the number of US scientists and engineers, although recovered after a drop, is about the same as a decade ago.

— US dominates in the number of science degrees. USSR and Japan dominate in the number of engineering degrees.

— US lead in publishing and patents has diminished. In earth and space science the percent of US technical articles dropped from 47% to 42% between 1973 and 1982. In engineering and technology the drop was from 42% to 38%.

Particular to aerospace:

— Published literature included in *IAA* is 50% non-US. The non-US percent is similar in the *Chemical Abstracts* and *Inspec* databases.

— More US aircraft patents are granted to foreign inventors than to US inventors.

We concur with the authors of the NSF study that a "nation's ability to use the results of scientific and technological activity conditions its ability to succeed in international political and economic competition". (16) If the doing of science is international, how well do we do in using the results? Not as well as we might.

True, the percent of references in US articles to articles from other countries is slowly increasing. In 1982 it was 46% in all S&T fields. It was 39% in earth and space sciences, and 44% in engineering and technology. However when we review a few key aerospace journals, we find that the citation rate for foreign literature is considerably less than warranted based on the origin of aerospace literature.

In the *AIAA Journal* references to foreign literature have remained only between 10—15% for 25 years. A spot check of the situation for two European aerospace journals, *La Recherche Aérospatiale*, and *Zeitschrift für Flugwissenschaften und Weltraumforschung (ZFW)*, shows a 40—50% non-European citation rate. Yet remember that 50% of the aerospace literature is non-US. NATO nations are not making full use of this wealth.

TABLE 1

Aerospace Literature Distribution vs Utilization

Literature Distribution (%)

US	Europe	USSR	Other
50	24	12	14

Literature Use — % Citations

	US	Europe	USSR	Other
AIAA J	85	11	1	3
ZFW	41	52	1	6
Aérospatiale	41	44	1	14

The percent of articles in major national journals by foreign authors is another measure of how one nation pays attention to the work of other countries. In US S&T journals 38% of the 1982 articles were by foreign authors. In earth and space science, 25% of articles were by foreign authors, and in engineering and technology, 32%.

The percent of articles coauthored by scientists and engineers from organizations in different nations is a last measure of international awareness. The NSF data measure this as a percent of all institutionally coauthored articles. In 1982 this was 17% in all fields, 31% in earth and space sciences, and 20% in science and technology. Looking closer to home, we identified the percent of all articles that were coauthored across international organizational boundaries. 2% of AIAA Journal articles were internationally coauthored; 4% of ZFW articles, and 6% of La Recherche Aérospatiale articles.

These data demonstrate that aerospace scientists and engineers do less well than other S&E in using the results of international research. We do have a noticable problem with barriers to international information transfer.

CONCLUSION

The barriers to international information transfer must come down. We have a big challenge in this conference to

determine how to do that. Some solutions are technological, such as electronic mail, machine-aided translation, standard, satellite transmitted education or conferences. Other solutions are nontechnological, including education in languages, in the value of information, and in the management and use of information resources. At AIAA we have a vision of the universally connected scientist/engineer who sits at a workstation which is connected to a gateway. The gateway supports the tools and resources to make those technological barriers fall. We believe that in the process of overcoming the technological barriers, the personal barriers will fall. I challenge you to get on with the conference, and on with the solutions. The seriousness of the situation tells us that there is no time to waste.

REFERENCES

1. — Strengthening US Engineering Through International Cooperation: Some Recommendations for Action. National Academy of Engineering. Washington, DC, p.1, 1987.

2. — ibid., p.3.

3. Wallerstein, Mitchel B. *US Participation in International S&T Cooperation: A Framework for Analysis*. In: Wallerstein, M.B., ed. Scientific and Technological Cooperation Among Industrialized Countries: The Role of the United States. National Academy Press, Washington, DC, 1984.

4. Greenberg, Daniel S. *Global Collider Research*. The New York Times, New York, p.A23. Sept. 8, 1987.

5. Fink, Donald E. *Lavi: A Tough Decision*. Aviation Week and Space Technology, p.15. Sept. 7, 1987.

6. Katz, Ralph
Allen, Thomas J. Investigating the Not Invented Here (NIH) Syndrome: A Look at Performance, Tenure, and Communication Patterns of 50 R&D Project Groups. R&D Management. 12(1): 7—19. 1982.

7. Lancaster, F.W.
Smith, L.C. Science, Scholarship and the Communication of Knowledge. Library Trends. 367—388. Winter 1988.

8. — National Science Board. Science Indicators, The 1985 Report. NSF. Washington, DC. 1985.

9 — NAE, loc. cit.

10. Michel, J. Linguistic and Political Barriers in the International Transfer of Information in Science and Technology. Journal of Information of Science. 131—135: 5(1982).

11. Van Bergeijk, D.
Risseeuw, M. The International Translations Centre: The Language Barrier in the Dissemination of Scientific Information and the Role of the ITC. Journal of Information Science. 37—42: 2(1980).

12. Allen, Lew Testimony before US House of Representatives, Feb. 4, 1987 in presenting *Balancing the National Interest: US National Security Export Controls and Global Economic Competition*. National Academy of Science. Washington, DC, 1987.

13. Lawrence, Barbara The Role of a Professional Society in Disseminiating Japanese Technical Information, in press. Information Services and Use.

14. — National Science Board. loc. cit.

15. Harford, James J.
Lawrence, B. Future Patterns of Scientific and Technological Information Exchange: A Scenario. AAAS Annual Meeting. New York City, May 1984 and current unpublished data.

16. — National Science Board. loc. cit.

LES BARRIERES LINGUISTIQUES ET CULTURELLES AU TRANSFERT DE L'INFORMATION

- LINGUISTIC AND CULTURAL BARRIERS TO THE TRANSFER OF INFORMATION -

par

Roland LARUE

Coordonnateur de l'Information Scientifique et Technique
du Commissariat à l'Energie Atomique,

Adjoint au Directeur du Centre d'Etudes Nucléaires de Saclay.

Centre d'Etudes Nucléaires de Saclay - Direction
91191 - GIF-SUR-YVETTE CEDEX

FRANCE

*
* *

RESUME

Un premier chapitre, décrivant l'importance relative des langues dans le monde en 1987, permet de se rendre compte que les habitants du globe fonctionnent avec une dizaine de langues, qui couvrent chacune une région du globe. L'analyse de la démographie et des variations des équilibres socio-économiques permet de dégager la tendance.

L'usage des langues est étudié, dans le deuxième chapitre, pour le domaine de la science et de la technique, en fonction du type de la communication (écrite ou orale) et du contenu (des mathématiques aux technologies) et les tendances sont dégagées.

Un troisième chapitre est consacré aux tentatives qui ont été faites pour réduire la barrière linguistique - comme le lancement d'une langue véhiculaire unique (esperanto, anglais) ou le développement de l'étude des langues étrangères. Les problèmes de traduction, de linguistique et d'ordinateurs ne sont pas traités.

L'analyse détaillée de la barrière culturelle, du quatrième chapitre, est effectuée par plusieurs approches (sociologique, psychosociologique, psychanalytique et psycholinguistique) afin de dégager les véritables enjeux individuels et collectifs.

Enfin, un dernier chapitre permet de dégager les faits porteurs d'avenir, comme le fonctionnement de la Commission des Communautés Européennes avec ses neuf langues officielles, et d'avancer quelques pronostics sur l'état linguistique du monde au début du 21ème siècle.

*
* *

SUMMARY

The first chapter, which describes the relative importance of languages in the world in 1987; shows that the worlds population effectively makes use of about ten languages, each of which covers a different area of the globe. The trends are revealed by an analysis of population density and socio-economic balances.

The use of languages in the field of Science and Technology is examined in the second chapter, with respect to the type of communication, either written or spoken, and the content, from mathematics to technology, and the trends are shown.

The third chapter is devoted to the attempts which have been made to overcome the linguistic barrier, such as the introduction of a single common language (Esperanto or English) or the development of the study of foreign languages. The problems of translation, linguistics and the computer, are not dealt with.

The detailed analysis of the cultural barrier contained in the fourth chapter is achieved by a combination of several different approaches (sociological, psychosociological, psychoanalytic and psycholinguistic) and reveals just what is at stake, both individually and collectively.

The final chapter concentrates on future trends, such as the working of the E.E.C. with its nine official languages, and makes several predictions as to the linguistic state of the world at the start of the twenty first century.

*
* *

SOMMAIRE - CONTENTS

*
* *

1 - L'IMPORTANCE RELATIVE DES LANGUES DANS LE MONDE EN 1987.

1.1 - La situation actuelle :

Depuis l'apparition du langage, les habitants du globe ont toujours communiqué entre eux par des langues locales, voire des dialectes, qui n'étaient réellement compréhensibles que par un nombre relativement faible de personnes constituant un groupe régional culturel.

Le vingtième siècle a vu la conjugaison de plusieurs phénomènes : explosion de la démographie, allongement de l'espérance de vie, accroissement du commerce et développement de la technologie, qui ont conduit les cinq milliards d'hommes vivant sur la terre en 1987, à communiquer entre eux à plusieurs niveaux.

Au niveau local, la langue - ou le dialecte - est toujours l'outil de communication du groupe culturel d'une région, avec cependant des caractéristiques nouvelles : le nombre d'habitants de certaines régions s'est beaucoup accru au cours du vingtième siècle (Algérie, Egypte, Indonésie, Pakistan...) et la bijectivité fréquente entre une langue culturelle et un état a conduit à un renforcement de l'appartenance à ce groupe culturel.

Au niveau global, les échanges en tout genre qui se développent depuis quelques dizaines d'années sur l'ensemble du monde ont été, de fait, précédés pendant deux ou trois siècles d'échanges importants mais limités à ce que l'on pourrait appeler les "grandes régions" du globe. Et comme il a bien fallu se comprendre pour commercer ou se battre, on constate qu'aujourd'hui onze langues sont, de fait, devenues "véhiculaires" pour l'une de ces régions :

- Mandarin	Chine
- Indi	Inde
- Anglais	Amérique Nord et Europe Occidentale
- Espagnol	Amérique Sud et Centre
- Arabe	Afrique Nord et Moyen-Orient
- Malais	Asie du Sud-Est
- Russe	Europe de l'Est
- Français	Europe Occidentale et Afrique
- Swahili	Afrique de l'Est
- Japonais	Japon
- Portugais	Amérique du Sud

1.2 - La tendance :

L'ensemble des grands équilibres démographiques, socioéconomiques et militaires ont toujours été déterminants pour l'importance d'une langue, aussi est-il difficile d'augurer de l'avenir. Cependant, si l'on exclut, pour le siècle prochain, l'hypothèse d'un grand cataclysme (biologique, militaire, géologique ou cosmique), le facteur démographique devient prépondérant et fait apparaître des modifications profondes dans l'importance des grandes langues du globe.

Alors les faits les plus marquants pour les cinquante ans à venir sont :

- la stabilisation de certaines langues : mandarin, indi, français, russe, japonais,
- la croissance du swahili, malais et portugais,
- la croissance rapide de l'espagnol

et la croissance encore plus rapide de l'arabe qui pourrait dépasser le mandarin au milieu du siècle prochain.

En quelque sorte la part des langues actuellement très importantes en nombre, comme le mandarin, l'indi et l'anglais va diminuer au profit essentiellement de l'espagnol et surtout de l'arabe.

Par ailleurs, les choix stratégiques politiques, comme la décision, en 1928, de Mustapha Kemal ATATURK d'imposer l'alphabet romain en Turquie et la décision de basculer sur l'alphabet cyrillique pour les esquimaux d'U.R.S.S., sont à la fois particulièrement importants au plan de l'exemplarité et de faible importance sur le plan quantitatif de l'évolution des langues pour l'ensemble du globe.

2 - *L'USAGE DES LANGUES EN 1987 DANS LE DOMAINE DES SCIENCES ET TECHNIQUES.*

2.1 - Nécessité de la communication scientifique internationale :

Pour un pays, les grands objectifs culturels de l'information scientifique et technique ont été synthétisés dans une circulaire (6) du ministre de la recherche français en septembre 1981 :

- exprimer l'identité scientifico-culturelle d'un peuple,
- assurer l'accès de ce peuple à la modernité, dans sa propre langue,
- assurer démocratiquement l'égalité de l'accès à la connaissance par le passage de l'information scientifique et technique dans l'ensemble du tissu social d'une nation.

Ce troisième point n'étant que la reprise d'un concept du philosophe ALAIN qui développait l'idée que "le niveau de culture d'un pays n'était pas défini par celui de Platon mais bien par celui de sa servante".

Pour un organisme de recherche, cette communication internationale est aussi nécessaire pour la défense de la qualité de l'organisme (son image) ou plus simplement pour la justification de l'utilisation des fonds alloués. Nous savons tous que les grands congrès périodiques internationaux sont en fait l'occasion donnée aux organismes de recherche de rendre compte de leurs travaux devant l'opinion publique.

Au niveau de l'individu enfin, la publication reste toujours nécessaire, malgré la réalité du "collège invisible" constitué par les personnes qui sont concernées par un même axe de recherche, car le problème d'antériorité se pose souvent, de même que, tout simplement, la consolidation du prestige personnel du chercheur ; il y a déjà plusieurs dizaines d'années, en effet, que le dilemme publish or perish est bien ancré dans la tête de tous les scientifiques.

En tout état de cause, c'est bien en science et en technologie que les investissements intellectuels, technologiques et financiers sont les plus grands et donc, globalement, c'est bien là que la nécessité de la meilleure information est la plus impérieuse. Par ailleurs, un rapport du D.O.E. (7) sur l'impact de l'Information Scientifique et Technique montre que l'investissement en I.S.T. est particulièrement rentable pour la recherche.

2.2 - Le type de la communication : orale, écrite :

Toutes les analyses (1, 2, 3, 4, 5, 6) font apparaître que l'anglais est, de fait, la langue véhiculaire la plus pratique pour la majorité des scientifiques.

C'est actuellement la langue la plus répandue quantitativement parmi les scientifiques et techniciens du monde entier, et elle est, de plus, située dans le groupe des langues qui, sur le plan de la qualité des publications sont les plus intéressantes.

L'anglais étant donc actuellement la langue véhiculaire de la communication scientifique, ceci amène une différence fondamentale entre le scientifique de langue anglaise et les autres.

En ce qui concerne la **communication orale** (3, 4), et particulièrement pendant les meetings, congrès ou conférences, les anglophones ont un énorme avantage naturel qui fait des non anglophones les grands perdants de la communication à tous les niveaux. Au niveau du message bien sûr, mais aussi et surtout à celui de la relation affective : la perte, dirait MAC LUHAN, est double, à la fois sur le "message" et sur le "massage".

Ce phénomène est encore amplifié par le fait que beaucoup de scientifiques anglais ou américains ne comprennent que l'anglais et qu'ils imposent en quelque sorte, à l'ensemble d'un groupe, le choix de la langue anglaise où ils sont bien entendu le plus à l'aise. Nous avons tous en mémoire des situations où un groupe discutant en allemand, français ou espagnol bascule sur l'anglais à l'arrivée d'une seule personne, au grand désavantage de tous les autres*.

---

* Je me souviens d'une réunion de négociation à Bruxelles réunissant 9 personnes de 6 pays : 8 originaires de 5 pays parlant couramment le français et une américaine ne parlant que l'anglais.

En ce qui concerne la **communication écrite** (3, 4), la situation est inversée et tourne à l'avantage des non anglophones.

En effet (3), 12 % des anglophones peuvent lire le français et 4 % l'allemand, alors que 97 % des francophones et des germanophones peuvent lire l'anglais. Ceci permet donc aux chercheurs non anglophones d'avoir accès à la littérature anglophone qui représente 50 % de la production des publications.

Ce résultat est corroboré par l'analyse du "Science Citation Index" qui montre que les Auteurs anglophones citent à 90 % des documents anglophones (alors que la production est de 50 %) et l'éventuelle différence de qualité ne peut pas expliquer une si grande différence, d'autant plus que deux phénomènes jouent en sens inverse : le fait que beaucoup des grandes revues internationales sont en langue anglaise est compensé par le fait que les documents achetés par les bibliothèques en langue autre que l'anglais sont beaucoup mieux sélectionnés (en raison de la difficulté d'accès) donc, dans doute, de meilleure qualité, et donc plus important à lire.

Un point assez remarquable de la communication écrite qui ne me paraît pas avoir été assez signalé est que certains spécialistes du Comité de lecture d'une grande revue internationale sont informés du contenu d'un article six mois avant les autres lecteurs. Tout se passe comme si l'information allait aux mieux informés.

2.3 - **Le type de contenu** :

A l'intérieur des sciences et des techniques, la situation linguistique n'est pas la même d'un domaine à l'autre.

Si nous prenons l'exemple des scientifiques français (3), nous constatons :

- que dans les domaines des mathématiques, de la physique et de la chimie, plus de la moitié des articles d'auteurs français sont rédigés en anglais et que ce pourcentage n'a fait que croître pendant les dernières années ;

- que dans les domaines des "sciences appliquées" (biologie, sciences de la terre, médecine et technologie), un article sur trois est rédigé en anglais, et que la croissance du nombre des publications en langue anglaise semble s'être arrêtée.

Par ailleurs, nous constatons que la part de la langue japonaise croît rapidement, particulièrement dans la technologie. Un rapport de 1986 (8) estime à 600 000 le nombre de publications scientifico-technologiques par an, dont moins de 3 000 font l'objet d'une traduction, même partielle.

Ces deux exemples autorisent le constat suivant : les publications en sciences fondamentales s'orientent vers l'écriture en anglais, celles des sciences appliquées sont essentiellement écrites en langue nationale.

2.4 - **La tendance** :

Comme l'information est plus vitale dans le domaine de la science et de la technologie que dans les autres domaines, l'abandon pur et simple du document en langue étrangère est moins fréquent, mais cependant encore alarmant. L'enquête sur les scientifiques anglais (2) a montré qu'en 1979, 29 % ne faisaient rien devant une référence française, alors que 84 % d'entre eux déclaraient se débrouiller en français (ce pourcentage d'inaction totale montait à 46 % dans le domaine des sciences sociales).

Par ailleurs, l'analyse des 500 000 rapports sur l'énergie détenus à Saclay, provenant de 45 pays, montre qu'ils ont été collectés pendant trente ans, parce que rédigés en langage dit "accessible", c'est-à-dire essentiellement en alphabet roman et cyrillique.

La règle semblait donc souvent l'abandon pur et simple du document "difficilement accessible".

Or, nous savons maintenant, grâce à l'enquête (7) portant sur 60 000 ingénieurs-chercheurs subventionnés par l'US.DOE qu'un "chercheur" lit en moyenne un document par jour de travail (220/an).

Ce chiffre de "un" document lu par jour et par individu a peu de chance de croître, alors que le nombre de publications annuelles augmente (8).

Le système étant manifestement en déséquilibre croissant, la solution de l'abandon de documents "difficiles d'accès" paraît de plus en plus risquée et des adaptations seront donc inéluctables.

On voit d'ailleurs apparaître ici que la quantité de documents sera en soi une grande difficulté, même si tous les documents étaient écrits dans la même langue. Il est probable que ce problème sera abordé par les deux bouts, à la lecture : sélectionner mieux les documents ; à l'écriture : écrire plus court et moins souvent*, puisqu'en cette matière les auteurs et les lecteurs sont les mêmes personnes.

---

* Un rapport du NTIS est lu en moyenne 9 fois, alors que certains (best sellers) sont vendus 50 000 fois... Il y en a donc quelques uns qui ont été écrits pour ... personne !

3 - LES TENTATIVES DE RESOLUTION DE LA BARRIERE LINGUISTIQUE.

3.1 - L'apprentissage des langues étrangères :

Les grands équilibres socio-économiques, politiques, religieux et militaires ont toujours déterminé le choix de l'étude d'une seconde langue.

Ainsi, pendant longtemps, pour être au courant de ce qui changeait dans le monde, une seule langue suffisait : le latin jusqu'au 16ème siècle et le français jusqu'au 19ème.

Ensuite, et jusqu'en 1950, il suffisait d'avoir la connaissance de trois langues romanes (français, allemand, anglais) pour être informé de la science et de la technologie, c'était donc encore possible.

Depuis, il a fallu ajouter une langue cyrillique : le russe, et aujourd'hui, une langue idéographique, le japonais.

Demain, il va falloir ajouter une autre langue idéographique, le chinois, et deux autres langues romanes, l'espagnol et le portugais ; et après demain, une langue arabique.

En conséquence, l'apprentissage de toutes les langues étrangères nécessaires à la communication scientifique n'est pas la bonne voie, d'autant plus que du côté de l'offre, la situation s'est beaucoup détériorée puisque seuls l'anglais et le français continuent à être enseignés dans la plupart des pays du monde, ce qui n'est pas de cas du russe, du chinois, du japonais et de l'arabe et plus le cas (6) de l'allemand et de l'espagnol.

Deux autres solutions seraient donc que tout le monde parle la même langue ou que tous les scientifiques apprennent une même "seconde langue scientifique" codée comme l'algèbre, et précise comme les systèmes d'unités.

3.2 - Une langue internationale unique.

La solution la plus radicale serait bien sûr la création d'une langue internationale qui éliminerait définitivement la barrière linguistique et c'est pour atteindre cet objectif que l'esperanto a été créé. Bien que l'opération "esperanto" soit importante* (5), elle est cependant restée à un stade de développement trop faible pour que l'investissement nécessaire à l'apprentissage de la langue soit rentable pour l'individu. Il faut dire que le développement de l'esperanto s'est trouvé concomitant de l'explosion démographique et pendant que quelques millions d'hommes apprenaient l'esperanto, quelques centaines de millions de chinois apprenaient "naturellement" le mandarin.

Le problème de la langue unique se pose un peu comme pour le téléphone ou la messagerie électronique : le grand nombre d'abonnés joue un rôle de première importance, la qualité est un peu le sous-produit de la quantité. La langue unique universelle reste donc fortement improbable dans un avenir proche, car elle nécessiterait l'apprentissage simultané de cette langue par deux ou trois milliards d'hommes.

3.3 - Une langue internationale pour la communication scientifique et technique.

L'anglais est la langue maternelle de 250 millions d'hommes et la deuxième langue pour également 250 millions d'hommes, mais ces derniers ne sont pas distribués au hasard parmi les activités des humains. Ils sont concentrés dans des domaines d'activité à haute densité de communication : essentiellement le commerce où 70 % du courrier international - surtout commercial - est en anglais, le domaine de la radio et de la télévision, le domaine des transports (aviation et marine) et le domaine de la recherche et du développement en science et en technologie.

Aujourd'hui 75 % des étudiants d'écoles secondaires du monde non anglais (à l'exception de la Chine) apprennent l'anglais.

L'anglais peut donc être considéré comme la langue véhiculaire de la communication scientifique et technique et, particulièrement dans le domaine scientifique où 50 % de la littérature est en anglais, on peut considérer que la barrière linguistique est dépassée, même si ce n'est peut-être que provisoire.

En effet, la quantité de scientifiques qui parlent vraiment l'anglais risque de se stabiliser assez rapidement car le nombre des étudiants qui iront dans des universités de langue anglaise (pour des études ou des recherches) ne pouvant croître indéfiniment, la proportion de ces réels bilingues va décroître.

4 - LA BARRIERE CULTURELLE.

4.1 - Les différentes approches.

Sans chercher à définir la culture, il est certain que la combinaison des aspects géographiques (terre, mer, plaine, montagne...) et des aspects climatiques (température, hygrométrie...) a été primordiale pour la constitution initiale des "groupes culturels" de notre espèce.

Ensuite, l'apparition d'une langue commune à un groupe a permis de transmettre la culture du groupe, mais aussi de structurer la société et d'en assurer ainsi la pérennité.

---

* Avec 100 journaux et 20 000 publications par an, l'esperanto compte quelques millions de pratiquants (1 à 8) et rassemble à son congrès annuel 4 500 personnes provenant de 50 pays.

La parole a amené une nouvelle répartition du pouvoir dans la société et l'apparition de l'écriture a conduit de nouveau à une nouvelle redistribution de ce pouvoir. - Le pharaon de l'époque avait bien raison de se mettre en colère et de s'arracher les cheveux lorsqu'on lui a présenté cette invention. -

Le développement technologique du vingtième siècle a eu un impact si considérable sur la vie culturelle des sociétés que l'on n'en mesure peut-être pas encore la portée. Si certaines avancées technologiques sont bien connues, d'autres sont peut-être aussi importantes sociologiquement : déverser du chlore dans les rivières est peut-être un facteur aussi important que la pénicilline pour l'espérance de vie ; les religions sont conduites à prendre en compte la synthèse des éléments par les étoiles et l'expansion de l'univers, et le nylon, par l'intermédiaire du collant qui a permis la mini-jupe, a sans doute participé de la libération de la femme dans certaines sociétés.

Psychologiquement, la fonction de la langue n'est pas uniquement de satisfaire le besoin de reconnaissance d'un invididu, mais aussi son besoin de considération. En fait la langue renforce le sentiment d'appartenance d'un individu à un groupe et inversement le besoin de connivence d'un groupe - pour se situer par rapport à d'autres groupes - développe la mobilité de la langue vers un dialecte ou même un argot qui, lui-même, change quand le groupe s'aperçoit qu'il diffuse dans les autres groupes.

Dès que l'approche de la barrière culturelle est un peu psychanalitique, on s'aperçoit du rôle prépondérant joué par l'inconscient. Deux enquêtes américaines rapportées dans "The foreign Language Barrier" (5) montrent que les deux facteurs déterminants pour exploiter la littérature soviétique étaient la connaissance familière de l'U.R.S.S. et la bonne opinion de la recherche en Union Soviétique, et ceux qui n'ont pas une bonne opinion de cette recherche ne citent jamais un auteur soviétique.

Sur le plan linguistique pur, si l'importance des paradigmes est facile à mettre en évidence (l'arabe possède plus d'une centaine de mots pour désigner le chameau alors qu'il n'y en a aucun pour désigner la table qui était un objet inconnu), l'axe syntagmatique est tout aussi important car il structure la pensée, parfois de façon exceptionnelle. - Certaines langues n'ayant pas la forme du futur (ou du passé), l'impact sur la pensée, la religion et la philosophie est considérable. -

Le cerveau de l'être humain qui, une fois adulte, fonctionne essentiellement par associations et par images (fantasmes) se développe dans la toute petite enfance à la fois par le langage et par la pensée ; c'est pendant cette phase de croissance que la pensée structure le langage mais aussi que le langage structure la pensée. Le résultat, bien sûr culturel, influence les approches scientifiques et technologiques. - La conception structurelle de certains ordinateurs japonais est imprégnée de culture japonaise.

4.2 - Les enjeux.

4.2.1 - L'enjeu culturel.

A chaque innovation technique, la quantité d'informations dont il a fallu disposer pour réaliser un objet s'est accrue. Par exemple, le coût de réalisation du film TRON - ou de l'ordinateur LISA - est de 50 M. $, soit 200 hommes X années.

Les investissements immatériels (intellectuels) vont devenir de plus en plus grands pour fabriquer les nouveaux produits, comme si la société de production allait se transformer en société de création.

La grande technologie du 21ème siècle sera celle de l'information sous toutes ses formes.

Chaque innovation technique a toujours, plus ou moins rapidement, modelé les rapports de forces entre les nations ainsi que les moeurs de ces mêmes nations. Ce qui fut particulièrement vrai pendant longtemps pour les armes, puis pour les biens de consommation, le devient maintenant pour les "biens de communication".

Il est symbolique que les deux premiers plus gros ordinateurs au monde (deux CRAY XMP) aient été achetés, l'un par le D.O.D. (Department de la Défense des U.S.A.), et l'autre par une petite entreprise de spectacle américaine ("Digital Production"), d'une trentaine de personnes, spécialisées dans la synthèse des images. Comme si le pouvoir et la prédominance des états se jouaient sur les deux fronts, celui de la guerre passive : les armes, et celui de la guerre active : la culture.

4.2.2 - L'enjeu de la langue.

Il ne faut pas rêver, une langue n'est ni neutre, ni innocente ; elle constitue, dans cette nouvelle forme de guerre économico-culturelle, l'un des enjeux les plus importants.

C'est sans doute pour cela que la langue commune à l'ensemble des habitants du globe a peu de chances de réussite dans un avenir proche.

De plus, cette solution unique n'est sans doute pas souhaitable car elle conduirait à trois risques :

- pauvreté culturelle généralisée par uniformisation,
- structuration unique de la forme de pensée,
- déliquescence de la langue choisie.

En outre, cela ne conduirait pas forcément ni à la paix mondiale (actuellement la plupart des belligérants dans le monde parlent la même langue), ni même à une meilleure efficacité dans le traitement des informations (le fonctionnement de la communication n'est pas plus performant dans les grandes entreprises multinationales).

5 - LES FAITS PORTEURS D'AVENIR SUR L'ETAT LINGUISTIQUE DU MONDE AU DEBUT DU 21ème SIECLE.

L'état linguistique du monde au 21ème siècle sera probablement le suivant :

- au niveau du "groupe d'appartenance", une langue "nationale" riche, socio-culturelle, porteuse de l'identité d'un peuple. Ces langues assurant les relations profondes entre les humains resteront très nombreuses ;

- au niveau régional, une petite dizaine - peut-être une demi-douzaine - de langues simples permettront la communication dans chaque grande région du globe ;

- au niveau mondial, une ou deux langues véhiculaires, surcodées mais pauvres, permettront d'assurer la communication "économique" globale.

Tout d'abord, quelques remarques s'imposent :

- pour faire aussi partie d'un autre groupe d'appartenance, il faudra apprendre réellement la langue de ce groupe ;

- pour avoir une action internationale, chacun devra faire l'effort d'apprendre la langue de la région et la langue internationale ;

- c'est toujours le besoin - ou l'envie - qui permet de sauter la barrière linguistique : la base PASCAL du C.N.R.S. est plus utilisée par le Japon que par les autres pays européens qui utilisent pourtant des langues romanes ;

- la communication dans les sciences appliquées - la "technologie" - se fera de plus en plus dans la langue nationale d'appartenance (8) parce qu'elle va concerner la totalité de la population du groupe, et peut-être aussi, dans certains cas, pour des raisons de confidentialité industrielle.

En attendant l'outil miraculeux qui permettra de s'exprimer vocalement dans sa langue d'appartenance, alors que l'interlocuteur entendra directement dans la sienne - et réciproquement - un certain nombre de choses peuvent être tentées pour réduire la barrière linguistique :

- en ce qui concerne la communication écrite, une voie pourrait être la traduction systématique, et en totalité, des revues en une dizaine de langues ; le premier pas vers cette traduction systématique pourrait concerner la traduction des résumés des revues secondaires afin qu'un document russe ou chinois puisse être retrouvé automatiquement, quelle que soit la langue dans laquelle le résumé a été traduit ;

- en ce qui concerne la communication orale, le premier objectif à atteindre est l'accoutumance de l'homme au multilinguisme (au moins deux), avec habitude de s'en servir, ceci étant particulièrement important pour les anglophones de langue maternelle anglaise. L'une des solutions les plus habiles (6) serait peut-être d'utiliser systématiquement deux langues pour toute intervention : la langue d'appartenance pour l'exposé et la langue régionale (ou internationale) pour les informations écrites projetées ; ou bien sûr l'inverse, suivant les circonstances et les possibilités linguistiques de l'intervenant.

Un second objectif à atteindre serait peut-être la prise de conscience de la dualité pureté/richesse, en effet, tous les "défenseurs" d'une langue "nationale" ont tendance à travailler "en défense", essentiellement en défense de la pureté de la langue, au détriment de sa richesse et de sa vie. Une expérience, comme ce que les sud-américains appellent le "portugnol" (mélange d'espagnol et de portugais et non alternance de phrases dans les deux langues) me paraît intéressante à suivre.

Une autre expérience à suivre est, évidemment, celle des organismes internationaux.

Ils fonctionnent généralement avec un petit nombre de langues officielles, deux pour l'Agence International de l'Energie (A.I.E.), quatre pour l'Agence Internationale pour l'Energie Atomique (A.I.E.A.), etc..., mais dans certains cas, le nombre des langues devient préoccupant. Ainsi la Communauté Economique Européenne (C.E.E.), qui regroupe douze pays, fonctionne avec neuf langues officielles, et consacre, en conséquence, plus de la moitié de ses dépenses de fonctionnement à l'activité traduction et interprétariat. Cette activité concerne plus du tiers des agents permanents de la Commission des Communautés Européennes (C.E.E.) et plus de la moitié de ceux du Parlement Européen !

Même si cette mécanique de traduction systématique en neuf langues peut apparaître comme monstrueuse - voire complètement débile - il me semble extrêmement important de suivre l'évolution de cette expérience car elle porte justement sur une dizaine de langues, c'est-à-dire le même nombre que celui des grandes langues régionales du globe. Ne serait-ce pas encore une fois une expérience d'avant - garde tentée par la "Vieille Europe", en avant-première et en vraie grandeur, de ce qui se passera dans le monde entier ?

* * *

BIBLIOGRAPHIE

(1) D.N. Wood. The foreign language problem facing scientists and technologists in the UK. J Documentation. 23(2) 117-130. 1967.

(2) S.K. Ellen. Survey of foreign language problems facing the research worker. Interlending review 7(2) 31-41. 1979.

(3) J. Michel. Director CDST/CNRS, Paris - France. Linguistic and Political Barriers in the International Transfer of information in science and technology. J Information Science, 5 131-135. 1982.

(4) J. Davies. UK. Comment on above article, stresses culture and group dynamics. J Information Science 6 179-181. 1983.

(5) J.A. Large. The Foreign Language Barrier, problems in scientific communication. Deutsch. 1983.

(6) J.P. Van Deth, F. Algardy, F. Rumpf. Parlez-vous Français ? Approches linguistiques de quelques congrès scientifiques internationaux. Rapport MIDIST/MIR. 1983.

(7) US.DOE Report - DOE/OR/11 232.1 (DE 820 14250). Value of the Energy Data Base. King Research, Inc. Rockville, Maryland. March 31, 1982.

(8) C.P.E. - Centre de Prospective et d'Evaluation. Bulletin n° 30, p. 29 à 32. 1986.

LINGUISTIC AND CULTURAL BARRIERS TO THE TRANSFER OF INFORMATION

by

Roland LARUE

Coordinator of the scientific and technical information of the Atomic Energy Commission,

Assistant to the director of the Nuclear Study Centre of Saclay.

Nuclear Study Centre of Saclay — Direction
91191 — Gif-sur Yvette CEDEX
FRANCE

SUMMARY

A first chapter, describing the relative importance of languages in the world in 1987, shows that the world population makes use of about ten languages, each of which covers an area of the globe. The trends are revealed by an analysis of population density and socio-economic balances.

The use of languages is studied, in the second chapter, in the field of science and Technology, according to the type of communication (written or spoken) and the contents (from mathematics to technology) and the tendencies are shown.

A third chapter is devoted to the attempts which have been made to overcome the linguistic barrier — such as the introduction of a single common language (esperanto, English) or the development of the study of foreign languages. The problems of translation, linguistics and the computer are not dealt with.

The detailed analysis of the cultural barrier, in the fourth chapter, is carried out by several approaches (sociological, psychosociological, psychoanalytic and psycholinguistic) in order to reveal the guenine individual and collective stakes.

Lastly, a last chapter concentrates on future trends, such as the working of the Commission of the European Communities with its nine official languages, and to put forward some predictions on the linguistic state of the world at the beginning of the 21st century.

CONTENTS

1 RELATIVE IMPORTANCE OF LANGUAGES IN THE WORLD IN 1987.

1.1 Current situation

Since language has appeared, the world population has always communicated between them by local languages, or even dialects which were only comprehensible by relatively few people setting up a cultural regional group.

The twentieth century saw the combination of several phenomena: boom of demography, lengthening of life expectancy, increase in trade and development of technology, which led the five thousand million people living on earth in 1987, to communicate at several levels.

At the local level, language — or dialect — remains the communication tool of the cultural group of an area, but with new characteristics: the number of inhabitants of certain areas has increased considerably during the twentieth century (Algeria, Egypt, Indonesia, Pakistan, ...) and the osmosis between cultural language and a State led to a membership strengthening of this cultural group.

At the overall level, exchanges of any kind which have been developed during the last decades throughout the world have been preceded in fact by important but limited exchanges in the "large areas" of the world. And as it had been first necessary to understand each other for trading or fighting, only eleven idioms became the sole communication means, for one of these areas:

Mandarin	China
Indi	India
English	North America and Western Europe
Spanish	South and central America
Arabic	North Africa and the Middle East
Malay	South-east Asia
Russian	Eastern Europe
French	Western Europe and Africa
Swahili	East Africa
Japanese	Japan
Portuguese	South America

1.2 Tendency

The whole of the major demographic, socio-economic and military balances was always determining for the importance of language, and so is it difficult to forecast the future. Yet, if, for the next century, the assumption of major cataclysm is excluded (biological, military, geological or cosmic), the outstanding factor is demography and it reveals deep changes in the importance of the major wordls languages.

Then the most outstanding facts for the fifty years to come are:
- Stabilization of certain languages: Mandarin, Indi, French, Russian, Japanese,
- Growth of Swahili, Malay and Portuguese,
- Rapid growth of Spanish and the even more rapid growth of Arabic which could be used more than Mandarin in the middle of the next century.

To some extent, the share of currently very important languages in number, like Mandarin, Indi and English will decrease to the profit of Spanish and above all Arabic.

In addition, political strategic choices, for example, Mustapha Kemal ATATURK's decision in 1928 to impose the Roman alphabet in Turkey or the decision to use the Cyrillic alphabet for USSR Eskimo are particularly important as far as exemplarity is concerned and of very limited importance as far as the quantitative level of languages evolution is involved.

2 THE USE OF LANGUAGES IN 1987 IN THE FIELD OF SCIENCES AND TECHNIQUES.

2.1 Need for international scientific communication:

For a country, the broad cultural aims of scientific and technical information were synthesized in a circular (6) from the French Minister of Research in September 1981:

- to express the scientific and cultural identity of people,
- to make modernity available to the people, in their own language,
- to ensure democratic equality to knowledge through the circulation of scientific and technical information within the social fabric of a nation.

This third idea being just a new formulation of ALAIN's philosophical concept when he stated that "the cultural level of a country is not determined by that of Plato but by that of his maid".

For a research organisation, this international communication is also necessary for promoting its quality (its image) or more

simply for justifying the use of the allocated funds. We all know that the major international periodic congresses give in fact the opportunity to the research organisations to publicize their work.

Finally at the individual level, being published is a must, despite the reality of the "invisible college" composed of people sharing the same research preoccupation, for the anteriority problem often arises, as well as merely the consolidation of the personal researcher's prestige; in fact, the publish or perish dilemma has been experienced for several decades and is deeply rooted in all scientists' heads.

In any event, science and technology are the two fields in which intellectual, technological and financial investments are the largest, and, as a whole, this is where the need for best information is most valuable. In addition, a DOE report (7) on the impact of scientific and technical information show that investment in STI is particularly profitable for research.

2.2 – Type of communication: spoken, written:

All the analyses (1, 2, 3, 4, 5, 6) reveal that English is, in fact, the most practical common language used by most scientists.

It is currently the most widespread language used among the scientists and technicians of the whole world, and it stands in the group of languages which, in the field of the publications quality, are the most interesting ones.

Since English is currently the common language of scientific communciation, this brings a fundamental difference between English language scientists and the others.

With regard to oral communication (3, 4), and particularly during meetings, congresses or conferences, English-speakers have the great advantage of the language over non English-speakers who are the great losers as far as communication at all levels is concerned, that means the message obviously, but more especially the level of emotional relation: the loss, would say MACLUHAN, is double, both on the "message" and the "massage".

This phenomenon is still amplified by the fact that many English or American scientists understand only English and that they impose to some extent, to the whole of a group, the choice of English language with which they feel at ease. We all have in mind cases in which a group discussing in German, French or Spanish start speaking English at the coming of only one English-speaker, to the major disadvantage of all the others*.

With regard to written communication (3, 4), the situation is reversed and turn to the advantage of non English-speakers.

Indeed (3), 12 % of English-speakers can read French and 4 % German, whereas 97 % of French-speakers and German-speakers can read English. This enables therefore non English-speaking scientists to access literature written in English which represents 50 % of the publications production.

This result is corroborated by the analysis of "Science Index Quotation" which shows that 90 % of documents written in English are quoted by English-speaking authors (for a production of 50 %) and the possible difference in quality cannot explain such difference, especially when one considers the two following opposite phenomena: Firstly, though most international reviews are in English, the documents bought by libraries in another language are better selected (owing to the difficulty of access) and therefore, undoubdedly, of a better quality and are more important to read.

A rather remarkable point of the written communication which has not been sufficiently pointed out in my opinion, is that certain specialists of the reading Committee of an important international review are informed of the contents of the article six month before the other readers, as if information was aimed at the best informed.

2.3 – Type of contents:

In sciences and techniques, the linguistic situation is not the same from one field to the other.

If we take the example of the French scientists (3), we note:

- in mathematics, physics and chemistry, more than half of the French authors' articles are written in English and this percentage has increased steadily for the last years;

— in the "applied sciences" (biology, earth sciences, medicine and technology), one article out of three is written in English, and the rise in publications written in English seems to to have stopped.

In addition, we note that the share of Japanese language rises quickly, particularly in technology. A report written in 1986 (8) gave an estimated number of 600 000 scientific and technologic publications a year, of which less than 3 000 are fully or partly translated.

These two examples justify the following report: publications in fundamental sciences are directed towards English writing, whereas publications in applied sciences are essentially written in national language.

---

* I remember a negotiation meeting in Brussels assembling 9 persons from 6 countries: 8 coming from 5 countries usually speaking French and one American speaking only English.

2.4 – Tendency:

As information is more vital in science and technology than in any other fields, pure and simple giving up of the document in a foreign language is less frequent, but is still alarming. A survey concerning English scientists (2) showed that in 1979, 29 % remained passive in front of a French reference, whereas 84 % stated that they had a working knowledge in French (this passivity reached 46 % in social sciences).

In addition, the analysis of the 500 000 reports on energy held in Saclay, coming from 45 countries, shows that they have been collected for thirty years, because written in so-called "accessible" language, i.e. primarily in Romance and Cyrillic alphabet.

So the rule often seemed to be the pure and simple giving up of the "difficult to reach" document.

But we do know now, owing to a survey (7) concerning 60 000 engineers-researchers subsidized by US.DOE that a "researcher" reads on average one document a day (220/year).

This figure of "one" document read a day and by individual has little chance to rise, whereas the number of annual publications increases (8).

As the unbalance of the system is obviously increasing, the solution of giving up the "difficult to reach" document appears more and more risky and the adaptations will be therefore inescapable.

It stands to reason that the quantity of documents will be a great difficulty, even if all documents were written in the same language. This problem will be probably tackled by both ends,
- at reading: to select the documents better;
- at writing: to write shorter and less often*, since in this matter authors and readers are the same people.

3 ATTEMPTS TO OVERCOME THE LINGUISTIC BARRIER.

3.1 Foreign languages study:

The major socio-economic, political, religious and military balances have always determined the choice of studying a second language.

For a long time, only one language was enough to be aware of what was going on in the world, Latin until the 16th century and French until 19th.

Then, and until 1950, three Romance languages (French, German, English) were enough to be informed of science and technology, it was therefore still possible.

Since, it has been necessary to add a Cyrillic language: Russian, and today, an ideogrammatic language, Japanese.

Tomorrow, it will be necessary to add another ideogrammatic language, Chinese, and two other Romance languages, Spanish and Portuguese; and after tomorrow, Arabic language.

So the study of all foreign languages to be used in scientific communication is not the right way, because concerning the offer, the situation has worsened very much since only English and French are still being taught in most of the countries of the world, but this is not true for Russian, Chinese, Japanese and Arabic and is not true any more (6) for German and Spanish.

Two other solutions would be therefore that everyone speaks the same language or that all scientists learn the same "second scientific language" coded as algebra, and precise as the unit systems.

3.2 A common international language.

The most radical solution would be of course, the creation of an international language which would eliminate the linguistic barrier definitively and it is to reach this aim than the esperanto was created. Although the operation "esperanto" has been of importance† (5), it has remained however at a development stage too weak so that the investment necessary to its study be profitable for the individual. It has to be said that the development of esperanto has been concomitant to the demographic boom and while a few million people learned esperanto, a few hundred millions Chinese learned Mandarin "naturally".

The problem of a common language is slightly the same as telephone or mail publishing: the large number of subscribers plays a major part, quality is a little the by-product of the quantity. A universal common language remains therefore strongly improbable in the close future, for it would require the simultaneous learning of this language by two or three thousand million men.

---

* A NTIS report is read on average 9 times, whereas some (best sellers) are sold 50 000 times .. There is therefore a few ones which were written for... nobody.

† With 100 newspapers and 20 000 publications a year, the esperanto is used by a few million people (1 to 8) and its annual Congress assembles 4 500 people coming from 50 countries.

3.3 International language for scientific and technical communication.

English is the mother tongue of 250 million men and is the second language of other 250 million men, but the latter are not distributed randmonly among human activities. They are concentrated in spheres of activity with high communication density: primarily trade in which 70 % of the international — mainly commercial — mail is in English, but also radio and television, transport (aviation and marine), and research and development in science and technology.

Today 75 % of the secondary school students in the non English world (except China) learn English.

Therefore English can be regarded as the common language inof scientific and technical communication and, particularly in the scientific field where 50 % of the literature is in English, one can consider that the linguistic barrier is a problem now out of date, even though this may be provisional.

Indeed, the quantity of scientists who really speak English is likely to stabilize rather quickly because as the number of students who will go in English language universities (for studies or research) will not grow indefinitely, the proportion of these real bilingual persons will decrease.

4 CULTURAL BARRIER.

4.1 Various approaches.

Without trying to define culture, it is quite clear that the combination of geographical aspects (earth, sea, plain, mountain, ..) and climatic aspects (temperature, hygrometry, ...) has played a major part in the initial constitution of the "cultural groups" of our species.

Then, the appearing of one language common to a group has allowed the culture of a group to be transmitted, but also society to be structured and thus perenniality to be ensured.

The speech has brought new distribution of the power in society and the appearing of writing led again to new redistribution of this power. — The pharaon of this time was well right to become angry and to tear his hair when this invention was presented.

The technological development of this century has so great an impact on peoples' cultural life that its effects cannot be ascertained yet. If certain technological projections are well-known, may be others are also sociologically important: to pour chlorine into the rivers is perhaps a factor as important as penicillin for life expectancy; the religions are led to take into account the synthesis of elements by stars and the expanding universe, and nylon, via sticking which allowed the mini-jupe, probably took part in the Women's Lberation in certain societies.

Psychologically, the function of language is not solely to satisfy the need of recognition of Man, but also his need of consideration. In fact language strengthen the feeling of belonging to a group and conversely the need of collusion with a group — to place in relation to other groups — favours language mobility towards a dialect or even slang which changes itself when the group realises that it spreads in the other groups.

As soon as the approach of the cultural barrier is somewhat psychoanalic, one realises the leading role played by the unconscious. Two American surveys reported in "The foreign Language Barrier" (5) show that both determining in exploiting Soviet literature were the familiar knowledge of U.R.S.S. and the good opinion of research in Soviet Union, and those who do not have any good opinion of this research never quote a Soviet author.

At the pure linguistic level, if the importance of paradigms is easy to demonstrate (Arabic has more than one hundred words to indicate "camel" whereas there is none to indicate "table" which was an unknown object), the syntagmatic function is quite as important for it structures thought, sometimes in particular instances. — For some languages which do not have the future (or the past), the impact on thought, religion and philosophy are considerable. —

The human brain which, once adult, mainly functions through associations and images (phantasms) develops in the very early childhood both by the language and thought; it is during this growth phase that thought structures the language but also the latter structures the former. The result of this process, which is of a cultural nature, influences the scientific and technological approaches. — The structural conception of a number of Japanese computers is impregnated with Japanese culture.

4.2 Stakes.

4.2.1 Cultural stake.

To each technical innovation, the quantity of information which has been necessary to carry out an object has increased. For example, the realization cost of the film TRON — or of the computer LISA — is 50 M.$, i.e. 200 men/years.

Immaterial namely intellectual investments will become more and more important to manufacture new products, as if the production society were being transformed into a creation society.

The major technology of the 21st century will be that of information in every aspect.

Each technical innovation has always, more or less quickly, determined the balance of power between nations and their way of life. This was particularly true for a long period of time with weapons, then with consumer goods. Now this is true for "communication goods".

It is symbolic that the first two biggest computers in the world (two CRAY XMP's) were bought, by the D.O.D. (Department of the Defence, U.S.A.) and by a small U.S. show business company ("Digital Production"), employing about thirty persons and specialized in image synthesis. Actually, it is as if the power and the prevalence of the states were at stake on two fronts, passive war: weapons, and active war: culture.

4.2.2 Stake of language.

There is no doubt about it: a language is neither neutral, nor innocent; within this new form of economic and cultural warfare, it it one of the most important stakes.

It is probably why the dream of a language common to everybody on earth has but a few chances of succeeding in the near future.

Additionally, this unique solution is probably not desirable since it entails three risks:

- cultural impoverishment due to uniformity,
- unique structuring of the way of thinking,
- degeneration of the selected language.

Moreover, this would not lead to world peace (at present most of the belligerents throughout the world speaks the same language), nor even to a better effectiveness in communication (the way of communicating is not more powerful in the major multinational corporations).

5 PROMISING FACTS ON THE LINGUISTIC SITUATION OF THE WORLD AT THE BEGINNING OF THE 21st CENTURY.

The linguistic condition of the world in the 2lth century will be as follows:

— at the level of the "membership group", "national", rich, sociocultural languages, representing the identity of the people. These languages ensuring the major relations between men will remain very numerous;

— at the regional level, about ten — perhaps six — simple languages to communicate within each large world area;

— at the world level, one or two overcoded but poor common languages, will allow information to be exchanged at the overall "economic" level.

But, before, a few remarks are necessary:

— to really participate in this membership group, one will have to learn the language of this group;

— to have an international action, each one will have to endeavour to learn both the regional and international language;

— it is always the need — or the wish — which makes it possible to overcome the linguistic barrier: the PASCAL base of C.N.R.S. is more used by Japan than by the other European countries which use however Romance languages;

— communication in applied sciences — " technology " — will be done more and more in the national membership language (8) because it willcreasingly concern the whole population of the group, and perhaps also, in some cases, for reason of industrial confidentiality.

Until comes the miraculous tool with which one expresses vocally in his membership language, whereas the interlocutor hears directly in his language — and reciprocally — a number of things can be attemped to reduce the linguistic barrier:

— with regard to written communication, a way could be the systematic — and full — translation of the reviews in about ten languages; the first step towards this systematic translation could concern the translation of secondary reviews' abstracts so that a Russian or Chinese document can be found automatically, whatever the language into which the summary was translated;

— with regard to oral communication, the first objective to be reached is the man's familiarization to multilingualism (at least two), and the habit of using it, this step being particularly important for English mother tongue speakers. May be one of the most skilful solutions (6) will be to use systematically two languages for any speech namely the membership language for talks and the regional language (or international) for planned written information; or of course the reverse, according to the circumstances and the speaker's linguistic performance.

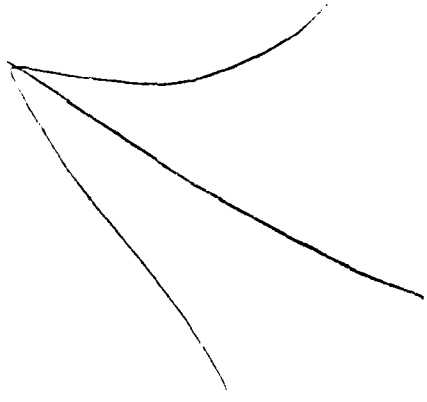
A second aim to be reached would be perhaps the awakening of duality purity/richness. Indeed all the "advocates" of a "national" language tend to defend essentially the purity of language, to the detriment of its richness and its life. In my opinion

an exeriment to be followed is that of the "Portugnol" — as South American name it — which is a mixture of Spanish and Portuguese with no sentences alteration in both languages.

Another experiment to be followed is, obviously, that of the international organizations.

They function generally with a small number of official languages, two for the International Energy Agency (A.I.E.), four for the International Agency for Atomic Energy (A.I.E.A.), etc..., but in some cases, the number of languages becomes alarming. Thus the European Economic Community (C.E.E.), which gathers twelve countries, functions with nine official languages, and devotes, accordingly, more half of its administrative expenditures to the translation activity interpretership. This activity concerns more third of the permanent agents of the Commission of European Communities and more than half of those of the European Parliament.

Even if this systematic translation procedure into nine languages can appear monstrous — or even completely weak — it seems to me very important to follow the evolution of this experiment for it covers precisely about ten languages, i.e. the same number as that of the major regional languages on earth. Would not it be once more an avant-garde full-scale experiment attempted by the "Old Europe", as a prefigurement of what will occur in the whole world ?



POLITICAL AND ECONOMIC BARRIERS TO INFORMATION TRANSFER

Jane Bortnick
Assistant Chief
Science Policy Research Division
Congressional Research Service
Library of Congress
Washington, D.C. 20540

SUMMARY

As international computer/communications networks proliferate, the issue of limits to information transfer becomes more critical. Individual businesses and even entire industries depend upon the ability to transfer information on a global basis in a timely and cost-effective basis. Additionally, international cooperation in science and technology involves greater international collaboration and joint efforts. Barriers that threaten this capability ultimately reduce the growth of markets and limit innovation. For this reason, efforts are underway in bilateral and multinational negotiations to reduce barriers and establish consensus guidelines in areas ranging from privacy to trade. This paper focuses on some of the specific limitations to information transfer, including privacy, trade restrictions, national security, and telecommunications regulations. It also highlights international efforts to reduce these barriers and establish common approaches to facilitate information transfer.

INTRODUCTION

The trend toward global endeavors in business, science, and the arts is facilitated by the growth of international computer/communications networks. Intracorporate networks allow for coordination of research, production, and marketing functions that increasingly are dispersed throughout the world. The services industries, such as banking and financial services, use computer/communications networks to transfer information instantaneously on a worldwide basis. (1) World markets have become vastly important to industries that previously served only domestic consumers. Similarly, the scientific community is taking advantage of the new information technologies. They use computers and communications networks for a wide range of activities. These applications include gathering data via computerized instruments, using computer models to simulate experiments, performing massive calculations using supercomputers, accessing databases via remote terminals, and exchanging scientific results through communication networks. (2)

Various political, legal, and economic barriers limit the full realization of these technologies and the benefits that accrue from their use. To a certain extent these barriers result from a world in which institutions have been unable to keep pace with the rapid advances in technological development. They also demonstrate the difficulties of making the transition to the "information age" in which information technology plays a critical role. In some instances, they represent attempts by governments to control the pace of technological change. In others, they reflect a desire to protect domestic industries from foreign competitors or to protect national security against adversaries. Different legal systems, along with diverse political and economic environments, also contribute to conflicting policies for information transfer between countries.

The growing recognition of the value of access to information has prompted an examination on both national and international levels of existing barriers and potential conflicts of laws and regulations. Efforts focus on harmonization of differing legal approaches to establish mutually agreeable guidelines for reducing barriers. Key to the success of these efforts is the development of a consensus on guidelines that provide for adherence to common principles, while retaining flexibility in their application.

PRIVACY PROTECTION

One of the earliest manifestations of the growth of computerized information systems was a concern for the protection of personal privacy. As electronic databases proliferated and remote access became commonplace, the need for effective safeguards for confidentiality of personal data emerged as a priority for many nations. In response to the potential threats posed by computer/communications systems, a number of nations enacted privacy or data protection laws. Most of the European laws covered only automated files, while in the United States privacy protection is linked to the individual rather than the medium in which personal information is stored. Generally the new data protection laws required registration of public and private databases containing personal information, as well as rights of access, notice, and correction. In some instances, these laws extend beyond individuals to cover "legal persons" (corporations or institutions) as well. (3)

In the late 1970s and early 1980s concern arose over the potential abuse of these laws to protect domestic industries, rather than strictly to provide individual privacy protection. Also, multinational corporations became alarmed over the necessity to adhere to differing privacy laws in different countries. There was a call for developing some consensus on basic privacy principles that would establish common ground rules, at least among the industrialized nations. The Organization for Economic Co-operation and Development (OECD) provided the forum for establishing agreement on these fundamental privacy principles. In 1980, the OECD Guidelines on the Protection of Privacy and the Transborder Flows of Personal Data were completed. (4) At the same time, the Council of Europe (COE) developed its own Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (5) for its member states. Both the OECD Guidelines and the COE Convention limit the collection of personal data, restrict disclosure to authorized purposes, and allow for individuals to inspect and challenge their records. The COE Convention, however, must be implemented through privacy legislation and may be applied to legal persons.

One of the key elements of the OECD Guidelines is the flexibility with which countries can demonstrate adherence to its principles. For example, in the United States there is no omnibus privacy law, but rather a tapestry of federal and state laws, which in conjunction with certain constitutional rights establish privacy protections. (6) Complementing this is the significant number of companies and other institutions that maintain internal privacy regulations. Establishment of these guidelines has eased concerns over both privacy protection and problems of operating in several different countries with different legal and regulatory regimes. Generally multinational corporations have indicated that privacy laws do not pose a major constraint in entering new markets and offering services. Although some concerns remain about the potential for privacy laws to limit operations, particularly in financial services, the development of consensus guidelines has proved beneficial for facilitating international information transfer.

ECONOMIC AND TRADE BARRIERS

Economic and trade barriers now pose greater threats to the transfer of information than privacy laws. As businesses and nations compete for greater portions of the global marketplace, the value of access to information becomes more critical. As indicated earlier, the successful operation of multinational enterprises is highly dependent on effective intracorporate networks. Further, information technology makes possible a growing array of new products and services that can expand current markets. No accurate figures exist for the value of world trade in services, but estimates exceed $360 billion and continue to grow at a substantial pace. (7) An important component of this is trade in telecommunications and computer services, not only because of its own growth potential, but also because it is critical to the operation of other service sectors.

Economic and trade barriers to these activities can range from high tariffs to incompatible standards. Telecommunications regulations often create another economic barrier to international computer/communications services. These barriers restrict not only the growth of the businesses providing computer/communications services, but also the activities of the users of these services. (8) A number of competing and conflicting interests arise in this area. How can intellectual property rights be secured, while simultaneously promoting the development of new technologies? How can effective standards for products and interconnection be established without inhibiting innovation? How can access to telecommunications facilities and services be liberalized while maintaining basic transmission capabilities? How can new trading rules be developed for services that will promote rather than discourage market expansion and openness?

Several examples may demonstrate how economic and trade barriers create obstacles to the transfer of information internationally. In some countries, there are restrictions on the use of foreign data processing facilities that require the use of redundant facilities domestically. There also may be requirements to purchase domestic technology, either by limiting imports or by restricting the type of equipment that can be attached to the public telecommunications network. Restrictive interconnection standards provide another barrier to the use of equipment that does not have approval of a country's telecommunications authority. Efforts to limit or deny access to flat-rate private leased lines create concerns as well for suppliers of international computer and information services. (9) Tariff structures in countries with telecommunications monopolies often provide a further disincentive to use of international computer/communications networks. Individually and collectively these economic and trade barriers limit expansion of information services and the technologies critical to delivering them. This harms not only the providers of computer/communications products and services, but perhaps more importantly their users.

Recently, there is evidence of a significant trend toward reducing the role of telecommunications monopolies and promoting competition in certain sectors. Traditional telecommunications monopolies are being replaced by a mixture of private and public entities providing a growing range of facilities and services. This reflects both technological developments and changing market structures. Today, more technological choices exist for transmitting information than ever before. Advances in satellites, as well as the spread of fiber optic cables, vastly increase available capacity. Lower circuit costs combined with tremendous reductions in computing costs make it possible to offer more services on a worldwide basis. The globalization of markets further contributes to the pressure to modernize telecommunications networks and provide innovative services.

The United States took the lead in promoting competition in telecommunications services. The development of competition in long-distance services, the divestiture of AT&T, the establishment of an "open-sky" policy allowing for domestic satellite competition, and the reduction of regulation by the Federal Communications Commission (FCC) all contributed to an environment in the United States which promoted greater diversity of equipment and services, along with lower prices. More recently, other nations have begun to revise their existing regulatory structure to allow for greater competition. In the United Kingdom, a second carrier (Mercury Ltd.) was authorized to compete with British Telecommunications, an independent regulatory entity was established, and competitive value-added services licensed. Japan also privatized certain telecommunications operations and now supports competition in various value-added services. (10) The French Government is in the process of separating the regulatory function from their telecommunications network supplier and allowing competition in value-added services. There is evidence that this trend toward greater competition in the provision of telecommunications services will continue in other countries. This should result in wider service offerings, lower prices, and greater innovation. The effectiveness with which the transition is accomplished will be significant for the user of telecommunications facilities and services and the ability to transfer information internationally. Currently, the United States is exerting considerable pressure on its major trading partners to ensure that as competition develops in these markets, there is equitable access for U.S. providers of computer/communications services.

NATIONAL SECURITY CONCERNS

Advances in science and technology are critical to the maintenance of national defense. As a result, there are legitimate concerns about access to our scientific and technical information that might compromise national security. However, in the United States the enormous advances in science and technology have resulted in part from an open exchange of information among scientists. The dilemma of how to strike an appropriate balance between openness and controls on scientific and technical information continues to create problems for policymakers. (11) In recent years, the problem has received significant attention for a number of reasons. The linkage between scientific developments and modern weapons systems is even stronger than before. The scientific enterprise is increasingly global in scale, supporting greater international collaboration. A number of laws, regulations, and Government policies have increased controls over the dissemination of technical information.

This has led to considerable public debate, in which the scientific community generally has spoken out in favor of reducing controls on the exchange of scientific information. They cite a number of incidents over the last few years where limits were placed on presenting papers at scientific meetings, foreign scientists were denied visas or not allowed to join collaborative research efforts, and prepublication review of scientific findings was requested. The Department of Defense (DOD) sought to clarify the issue by developing guidelines for publication of DOD-sponsored research. They stated that unclassified "fundamental" research performed at universities and Federal laboratories would not be restricted. Fundamental research is defined on the basis of the performer, budget category, sponsoring DOD entity, and special contractual provisions. (12)

Most recently, however, the information industry as well as the users of scientific and technical information, became concerned about the creation of a new "sensitive but unclassified" category of information. In 1984, President Reagan signed National Security Decision Directive (NSDD)-145 to improve security of Government computer systems. As part of this effort, a new National Policy on Protection of Sensitive, But Unclassified Information in Federal Government Telecommunications and Automated Information Systems (NTISSP No. 2) was issued on October 29, 1986. This policy statement had the potential for limiting items to be entered into databases controlled both by the Government and the private sector and raised questions about limiting access to what had previously been publicly available information. The new initiative received considerable criticism in congressional hearings (13) and subsequently was rescinded by the National Security Advisor.

The National Academy of Sciences (NAS) also explored extensively the issue of controls over disseminating scientific and technical information and technology transfer. In 1982, NAS issued the report Scientific Communications and National Security in which they found that the Soviet Union acquires a substantial amount of U.S. technology relevant to military systems. They further found, however, that very little of this transfer came from universities and open scientific communication. (14) In early 1987, a second NAS panel established to assess the effect of controls on commercial activities in high technology issued its report, Balancing the National Interest: U.S. National Security Export Controls and Global Economic Competition. The panel found that it is as important to maintain the Western technology lead through open scientific communication and trade, as it is to control the technology flow to the Soviet bloc. (15) As the NAS studies indicate, this is an extremely difficult question which requires ongoing assessment to ensure that neither national security nor scientific advancement are jeopardized as information transfer policies are developed and implemented.

EFFORTS TO REDUCE BARRIERS

There is a growing recognition worldwide of the importance of facilitating information flows to promote scientific advancement and commercial developments. A number of nations, both on a bilateral and multilateral basis, are working to establish agreements to reduce barriers to information flows. In 1985, the OECD Council of Ministers adopted a Declaration on Transborder Data Flows. The Declaration promotes cooperation on access to information and the avoidance of barriers to the international exchange of data. It also seeks to make regulations transparent and to develop common approaches for dealing with transborder data flow issues. (16) The Declaration also serves as the basis for future work by the OECD in facilitating transborder data flows.

A major effort is underway in the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) to establish rules for trade in services similar to those for trade in goods, as well as to improve intellectual property rights internationally. The results of these negotiations are bound to have a significant impact on the provision of computer/communications services internationally. In the same timeframe, the World Administrative Telephone and Telegraph Conference (WATTC-88) of the International Telecommunications Union (ITU) will establish definitions that will form the basis of regulations for international telecommunications services. The expansiveness of these definitions will be significant for determining what computer/communications services fall within the purview of international telecommunications regulations. As indicated earlier, the ability to provide competitive services on a cost-effective basis is important for both suppliers and users of information technology. Both of these multinational negotiations will establish an international regulatory framework that may either promote or inhibit international information transfer. For these reasons, considerable time and resources are being expended by the U.S. Government and the private sector to work toward a successful outcome of these negotiations.

REFERENCES

1. Feketekuty, Geza and Kathryn Hauser, "A Trade Perspective of International Telecommunications Issues", Telematics and Informatics, Vol. 1, No. 4, 1984, p. 359-369.

2. U.S. Congress. House of Representatives. Committee on Science and Technology. Task Force on Science Policy. Prepared by the Congressional Research Service, Library of Congress. The Impact of Information Technology on Science, 1986, Background Report No. 5, p. 7-10.

3. Pipe, G. Russell, "International Information Policy: Evolution of Transborder Data Flow Issues", Telematics and Informatics, Vol. 1, No. 4, 1984, p. 409-413.

4. OECD Guidelines on Privacy Protection and Transborder Data Flows, Organization for Economic Co-operation and Development, 1980, Press/A(80)57.

5. Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Council of Europe, 1980.

6. Aldrich, Robert, U.S. Department of Commerce, National Telecommunications and Information Administration, Privacy Protection in the United States, 1982, NTIA Report 82-98.

7. Sauvant, Karl P., International Transactions in Services: The Politics of Transborder Data Flows, Boulder and London, Westview Press, 1986, p. 1-17.

8. Robinson, Peter, "Telecommunications, Trade and TDF", Telecommunications Policy, December 1985, p. 310-311.

9. Office of the United States Trade Representative, U.S. National Study on Trade in Services: A Submission by the United States Government to the General Agreement on Tariffs and Trade, December 1983, p. 183-193.

10. Bruce, Robert R. From Telecommunications to Electronic Services, Boston, Butterworths, 1986.

11. Relyea, Harold C., editor, Striking a Balance: National Security and Scientific Freedom First Discussions, American Association for the Advancement of Science, 1985, p. 1-5.

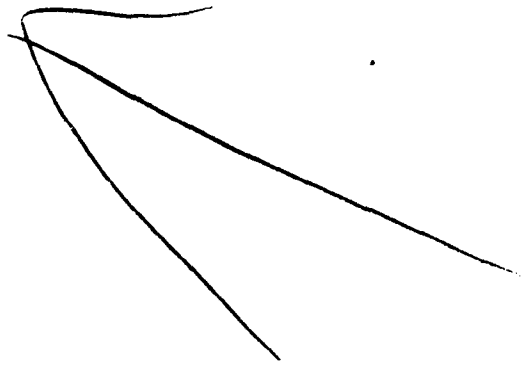
12. DeLauer, Richard D. Undersecretary for Research and Engineering, U.S. Department of Defense. Publication of the Results of DOD Sponsored Fundamental Research, Oct. 1, 1984.

13. U.S. Congress, House of Representatives, Report to Accompany H.R. 145, Computer Security Act of 1987, June 11, 1987, Rept. 100-153, Part 1, p. 17-20.

14. National Academy of Sciences. Committee on Science, Engineering, and Public Policy. Panel on Scientific Communication and National Security, Scientific Communication and National Security, Washington, National Academy Press, 1982, 118 p.

15. National Academy of Sciences. Committee on Science, Engineering, and Public Policy, Panel on the Impact of National Security Controls on International Technology Transfer, Balancing the National Interest: U.S. National Security Export Controls and Global Economic Competition, Washington, National Academy Press, 1987, p. 19.

16. Declaration on Transborder Data Flows, Organization for Economic Co-Operation and Development, 1985, Press/A(85)30.



Linguistic-Technical Aspects of Machine Translation

Harald H. Zimmermann
Universität des Saarlandes
5.5 Informationswissenschaft
D-6600 Saarbrücken 11
FRG

Summary

To allow to compare computer aided translation (CAT) and machine translation (MT) systems, essential criteria and typical exponents of the various concepts are presented. Among the most important criteria of differentiation are the following:

- Translation of complete texts (including titles, abstracts) vs lexical aid for translation;
- All-round translation vs translation of specialized fields and texts;
- Bilingual translation vs multilingual translation;
- User-oriented vs administrator-oriented expansion of dictionaries and systems; and
- Portability.

Included essential criteria for evaluation are:

- Quality of the raw (also called informative) translation; - advantages / disadvantages as opposed to human translation,
- Embedding in a user environment (data bases, indexing and retrieval, word processing, electronic publishing);
- User friendliness
- Costs of development and installation; and
- Running costs and effects of realization.

List of symbols

CAT	Computer aided Translation
EC	European Community
EUROTRA	European Translation System, research project of the European Community
LOGOS	Commercially available Machine Translation System of LOGOS Corporation
MINITEL	Videotex system access tool to the French TELETEL
MT	Machine Translation
PC	Personal Computer
SUSY/STS	Saarbrücken Translation System / Service
SYSTRAN	Commercially available Machine Translation System of World Translation Center, U.S.A., and Gachot S.A., France
Wang OIS	Office Information System of Wang Computers
WordPerfect	Commercially available editor

Introduction

Considering the technical development in information industry, one has to realize that the computer will be integrated in nearly every professional human activity. In the field of word or text processing, man-machine interaction, on PC level or on the base of a work station on terminal instead of a typewriter, will be the standard in the near future, that means from the end of the eighties.

On the other side, the problem of automatic translation of natural text in general - even not to speak about speech understanding and speech translation - will not be solved in the sense of HQFT (high quality fully automatic machine translation), due to the complexity of natural languages. So there are limits (more or less in the scientific approach) and possibilities of integrating technical tools in the process of (human) translation or even text understanding by machines.

The aim of this contribution is to explore the possibilities of usage of the computer in the field of machine and machine aided translation. Even if it might be of some attraction to treat also principal problems of translatability of texts (and understanding), we will concentrate on technical translation, that means the translation of technical or common texts.

Systematic aspects

Even if it might be of some interest how machine translation is realized, the linguistic aspects (esp. the models of grammar and the possible or used strategies) play an subordinate part. So one can assume that there is a kind of "black box", where a text or words in one natural language are put in and, with or without human interaction one or alternative translations come out of the system. The translation might be good, useful or bad under the aspects of the user.

Concentrating on the usage and usefulness of machine translation or computer aided translation, one has - in principle - to distinguish between two main user groups:

(1) the so-called end user, e.g. an expert trying to get the information of an article written in a natural language more or less unknown to him; a person writing a letter to a friend in a foreign language ... and

(2) a professional agent, esp. a human translator who tries to use the machine as a tool in the process of fulfilling his job.

Under these aspects, one can let aside projects and basic re search on the principles of machine translation / language understanding and orientate on practical tools resp. systems.

As a starting point, it is important to distinguish between two main strategies: machine translation (MT) and computer aided translation (CAT).

- A system will be called a machine translation system (MT) only if a translation process - starting from the machine readable source text - is fulfilled without any human interaction to reach a target text quality at least "good enough" for information purposes.

- A system will be called a computer aided translation system (CAT) if a human interaction is needed or foreseen to reach the aim of a "good" translation of a source text (machine readable or not).

It is quite clear that a MT system can be used, in addition, as a component of a CAT: a text can be adjusted before the MT starts to get better machine translations ("pre-editing"), and/or a machine translated text can be post-edited by human translators to get higher quality.

There are a lot of systems on MT and CAT on the market claiming to be the right practical tool, and the choice is hard to be made without having precise criteria for the decision.

Because the amount of data in computerized dictionaries - on the long line of development of MT and CAT systems - is the decisive component, the update of the (electronic) dictionary plays a substantial part in both alternatives:

- One concept may be - especially if the system still has a large capacity of dictionary entries - that the user, e.g. the translator should not take part at such a process (see, for example the SYSTRAN-concept), so that there are specialists needed on the system administration side to do the job of improving the dictionary data base;

- the other way is to let it to the professional user to complete the system's dictionary or to add a special vocabulary (see, as an example, the LOGOS-concept).

There are other system aspects which play an important part on the decision process: the availability of language pairs, the possibility of using or handling special text types (e.g. minutes, letters, ...).

Criteria of evaluation

Quality

Even if one normally can (or has to) handle the system as a "black box", there are differences in the quality of the "pure" machine (raw) translation results (output). It is not very easy to give a precise measure, but there are some important criteria to be noted (see, for details, esp. the concept of Van SLYPE (1982) and the descriptions of the two SYSTRAN evaluations, Van SLYPE (1979)).

The main criteria under these aspects are:

(1) reliability and fidelity, that means: to what degree (bad, good enough, good) the content / meaning of the original text is conserved.
(2) understandability and intellegibility, that means: to what degree the end user is able to read and understand the translated text.

This looks quite simple, but the problems occur in the detail: So, on the syntax and stylistic level, a system normally will produce translations which are worse than human translations, whereas on the lexical level, esp. in the identification of the right technical term(s), a system's translation may even be more precise and consistant than a human translation.

Application environment

A MT or CAT has to be seen in the (technical) environment of application possibilities. So, normally a decision is not only quality oriented, but also based on the possibilities of integration into a complex text or word processing system. Under these aspects, the following components have to be considered·

Integration in (bibliographical or textual) data bases

Meanwhile, data bases are - technically spoken - world-wide accessible via packet switching networks and even satelite communication. So, overcoming the language barriers, e.g. between English and Japanese, but especially in the European multilingual market, becomes an important desire. Experimental efforts to integrate MT systems in such an information process are made in Japan: as one example the INSPEC data base which is originally English can be accessed with Japanese key words. The key words and later the (English) title are automatically translated during the dialogue into Japanese (Nagao et al. 1982). In a similar way, a MT system for German to English is used via a batch process and with post-editing) to translate the titles of German data bases (Zimmermann et al 1987).

It is quite clear that translation of titles and abstracts could also be done by human translators. But there are some arguments for MT and CAT: the text to be translated is machine readable, so there is an ideal base for using a computer, the fields or areas on which the title / text is oriented are normally "physically" marked, so that the classification or even thesaurus functions can be used especially for the lexical transfer (disambiguation), the technical vocabulary needs to be very precise, so that the computer helps on beeing consistent.

Automatic Indexing

Indexing of (full) text might be a good side-effect of using MT and CAT. For lexical transfer, one needs to derive word forms to basic forms; compounds and complex words have to be identified as such, word class information, even relations between terms are used for disambiguation purposes. So it can be considered to provide such output or intermediate results of MT systems for the purpose of document archiving and information retrieval.

Text and word processing

There is no doubt that text processing plays an important part in every translation environment. Even free-lance (human) translators more and more will use a word processing system (on PC), and there is a small step to integrate, in any way, (private) glossaries or word lists accessible via so-called "windows" on the screen - instead of using card-index boxes. It is quite clear that other functions, e.g. spelling, grammar and style checkers, will increasingly be integrated in such a process.

As a result, the "source text" is machine readable. But the MT and CAT system has to be adapted to the (different) word processing systems (for example the Wang OIS is combined with LOGOS and SYSTRAN, WordPerfect is combined with SYSTRAN). If such tools are available, post-editing of MT results can be supported by the special editor.

One problem in this environment is the combination of those tools with local facilities of MT (see, e.g., LOGOS) or the connection with a Translation Service Center (see, e.g., the concept of SYSTRAN application in the European Commission or even the use of the MINITEL-System in France to get machine translations (by SYSTRAN).

There is no doubt that the usage of MT and CAT will make progress mainly in combination with text processing and online access. The question is, for the moment, if the existing tools are powerful enough (in quality) that they will be accepted by the user. Especially the results of the MT experiment done by Gachot S.A. on MINITEL and on PC-Access (with SYSTRAN) will be an important contribution in this direction.

Electronic Publishing

More and more, texts to be translated (esp. technical text like repair instructions, manuals) are fully prepared - including figures, drawings, tables, pictures - via electronic publishing, nowadays also known under the variant of desk top publishing. Companies giving commissions to translators - inside or outside the firm - don't want to rearrange (or compose) the complete product of the translation for even a lot of target languages.

So a great effort has to be undertaken (may be, on both sides: the producer of electronic publishing systems and the producer of MT and / or CAT software tools) to integrate translation helps without deleting or violate the document structure. Of course, there are problems in line, paragraph or page adjustment (due to the different length of translated text), and also the rearrangement of phrases / words due to the different word order raises problems on the correct integration of typographical markers (bold face, underlining etc.) in the target language. If these "technical" components of translation processes will not to be let to the human translator / posteditor, a higher standardized level text description must be integrated in the electronic publishing facilities. This will be a great challenge to the existing and coming MT and CAT systems.

User friendliness

On behalf of the professional translator's work station, the user friendliness plays an important part on every "level" of MT and CAT. Nobody should want that in the use of MT or CAT the human actor has to play the part of a "slave", e.g., correcting, day by day, only the "trivial" mistakes of the system's output. There is, at the moment, such a danger, because the existing systems are not very flexible and adaptive.

Future developments in MT and CAT must therefore concentrate on activities which give the user more and direct feedback possibilities. The "private file concept" which sometimes is used in data base systems, where a user can select and create a "personal part" in a data base, could be an example: at least on the dictionary level the user should get functions to realize - on the basis of the existing data - his "own" dictionaries (physically or logically).

By the way, both sides - the provider of the data base system and the user, can make profit of such a concept: the system's specialized vocabulary will be ameliorated and the user will have a great (but also responsible) influence in the choice of the translations.

What is true for the lexical part, should also be applicable to structural components (e.g. to influence the length of sentences, some stylistic components etc). Existing systems should get more flexibility and coming systems should consider such components from the beginning.

Cost and benefit

As everywhere in business, the decision of using a tool like MT or CAT is made on a cost and benefit calculation. In this case, it is not only a pure monetary problem, because "time is money" and getting translation without delay might be paid on a higher price.

But at the end, the decision is done on an economic base, having in mind the social or human effects.

Because there is no substantial data available, at least to the author, on the cost of development of MT and CAT systems, one has to concentrate at the cost and benefit on the user's side.

It seems, at the moment, that the amount of translations (measured in pages / day) which are produced by a human translator via interaction and / or post-editing can be substantially increased. On assuming that, for example, the maintenance and technical application of a system like SYSTRAN will cost $300.000/year, and that 300.000 pages can be technically translated by the system, the cost of the pure translation - not including the update of the dictionaries and the preparing or post-editing of the MT-results - nearly can be neglected ($1 / page). The cost of the complete process (translation with man-machine interacting) differ depending on the quality one wants to get. For a "good-enough" translation - which means informative translation, say, for example, for working papers, a so-called rapid post-editing can be done, so that a human translator produces about 20 pages / day (instead of 4-6 pages without any MT).

To reach high quality comparable to professional human translations, one has to consider more time for revising or post-editing. But it seems, that the break-even point is reached in the sense that if the vocabulary of a MT system is adapted to the user's field, the cost of translation are less than the cost of pure human translation, even considering that the use of word or text processing systems in the translation process spares about 20 % of time.

Basic linguistic an strategic problems

The morphological component (that means problems of inflection, derivation, and decomposition) in MT has been successfully solved, at least for practical purposes in the application environment, even if the correct translation of identified derived and decomposed words not always is reached via automated rules. This is not the case for solutions on the syntactic or semantic level. Even if we assume that a problem, e.g. the disambiguation of syntactic homographs (like RAINS in IT RAINS or THE RAINS) can be solved via a strict and fully formalized parsing system, the complexity of natural language structure leads to an explosion of computer time if one tries to integrate or handle every possible (partial) structure or occurence. So, in reality, commercially available systems try to shorten the process of identification (or disambiguation) via special deterministic rules or probabilities. As a result, they run 10000 or 1000 times faster than a fully linguistic oriented system, but their results may not reach the same quality.

Today, computer time plays not any more the same role as some years ago, but in machine translation computer time up to now is not fully neglectable. The same is true for the solution of problems of homonymy, that means in the semantic field. On the one side, there are limits in the handling of text structure vs. sentence structure. In most cases, the knowledge base of a MT system is the sentence environment, that means that information or solutions of previous sentences are lost and that nearly no data is known at the text level. This leads to many problems, especially in the field of pronominal reference, but also in article insertion and homonym disambiguation. The way existing systems are handling the general problem of semantic ambiguities is by introducing semantic codes (which - on a general level - also play some role in disambiguation of syntactic structures), especially "field" or discipline markers, which are used to select the "right" word (or even word sequence) depending on field parameters given by the user. They also try to solve this problem by dictionary look up to identify word sequences or even sayings (which normally have to be lexicalized because there exists no algorithmic solution, see the German ES REGNET BINDFADEN, which has to be translated by IT RAINS CATS AND DOGS.

Since Chomsky, the systematic structural-semantic access to language analysis, "understanding" and translation has made some progress. So, on the research level, there are several modern formalized grammar types and parsing systems available. Especially in Japan (see, e.g., the MU-System) and Europe (e.g. the efforts made by the European Community and their member states with the European Translation System, EUROTRA) research in MT is continuing. But it seems that one needs more than computational linguistic development: linguists, computer specialists, information scientists and users have to cooperate in large-scale projects to reach the aim of practical usability.

Examples

To give some impressions on the state of the art of so-called productive (not to say commercial) systems, and also to show how the mentioned criteria can be applied, two systems, the MT systems SYSTRAN and SUSY/STS will be described.

SYSTRAN

SYSTRAN (commercial rights at Gachot S.A., France) - in its newest version 3.7 - has the following characteristics:

- Translation of full text. Even if the structure is not correct or words are misspelled or words are not found in the machine dictionary, a translation is done.

- The speed of translation (depending on computer capacity) reaches up to 350.000 words per hour. So it is the fastest system available on the market.

- The system is language-pair oriented. Translations are available for the pairs English -> French, English -> Italian, French -> English, Russian -> English (USAF), English -> Japanese (SYSTRAN JAPAN), English -> Arabic, under development are, besides others, English -> German, French -> German, German -> English and German -> French. The quality depends, on the one side, on the availability of (discipline oriented) dictionaries. A great effort has been made at the European Community to develop the SYSTRAN dictionaries. For the translation from English to French, the dictionary now contains more than 150,000 entries. The same quality is not reached, e.g., for translation from German to French, which is in the starting phase, even on the level of language analysis.

- SYSTRAN application needs technical specialists (and administration). So only companies which are able to realize a special staff (e.g. the EC or USAF) have the possibility to use a SYSTRAN version on their own computer (if the computer is a mainframe IBM or IBM compatible). But there is an interesting alternative: to use the system via telecommunication networks, e.g. packet switching or - as a very "futuristic" example - via videotex. In France, a videotex application (using the French videotex version, called TELETEL, via a telephone combined with a screen, called MINITEL) is already available (and even used by pupils).

- Dictionary maintenance - up to now - for SYSTRAN dictionaries normally has to be done by system experts. The main problem is not the coding itself (which is very complex, but could and will be handled by user-friendly interfaces), but the consistency of the dictionary data base. Dictionaries contain area codes, but this component has to be developed to get a more flexible user and usage orientation.

- SYSTRAN is nearly not portable, that means: the basic code is an IBM-Assembler, even if the linguistic rules normally are written in a special macro language. The system itself needs - as mentioned above - a mainframe (IBM or SIEMENS or AMDAHL) computer or computers of similar size. A software revision (may be, on the base of UNIX) is planned.

- The quality of (raw) translation differs depending on the language pairs and dictionary satisfaction: If one concentrates on English -> French, the following percentages may apply: Morphological identification: about 100 %; syntactic structures: about 90 %; semantic disambiguation: between 80 and 90 %, depending on the integration of so-called limited-semantic rules.

- SYSTRAN can be combined with several word processing environments. One is the Wang-OIS (which is used at the EC), but one can also use PC (IBM compatible) with an editor like WordPerfect. There are applications where one uses an optical character reader to create a text file with WordPerfect, sending the data to the SYSTRAN service center (e.g. at Gachot S.A., Paris) and getting back the translated text via post line and a software tool which allows to post-edit the text having both versions - the original and the translation - on a splitted screen.

- The usage of the translation system itself is batch oriented. During the process of translation, there is no possibility of intervention (despite the fact that the system administrator has the possibility to introduce translations for unknown words). On the other side, the user handles the system as a black box, he doesn't need any knowledge of the system.

- There is no information available about the development cost of the system. One can estimate that they are between $ 20,000,000 and $ 50,000,000. The EC application alone cost about $ 4,000,000 and $ 5,000,000. The cost running a version (without royalties) lie - in my opinion - at $ 300,000 / year. So one can estimate that if one is able to translate about 300,000 pages a year (e.g. at the EC), that the system

cost are less than $ 1 / page (word processing, post-editing and computer time not included). The translation fee on the French MINITEL (vidsotex) systems is about $ 0.10 per "window" (up to 10 lines), the automatic (raw) translation normally is available in this application within 45 to 60 seconds.

- This leads to the conclusion that SYSTRAN is end-user oriented, that means that its market is the "information society" where 100 % quality is not needed, but good-enough translation to understand a text written in a language not or nearly not known by the user. But there is no doubt that SYSTRAN has also a good chance to be used in the professional translation environment as a high speed tool and a possible alternative in computer aided translation.

SUSY/STS

SUSY/STS, developed in the framework of a large research project in computational linguistics at the university of Saarbrücken (FRG), now is used as production system in the field of translation of data bases. This service is done by the Institute of applied Information Research (IAI) at Saarbrücken. The main characteristics are:

- Translation of full texts (sentence level) is possible. In the application environment, one concentrates on title (and abstract) translation.

- As SYSTRAN, SUSY/STS is an "all-round" and robust translation system: A text is translated even if there are unknown words included. In such a case, the original word is integrated in the text of the target language.

- To some extent, SUSY/STS is a "multilingual" MT system. Only the transfer component (between two languages) is bilingual, whereas the analysis and synthesis is done independent from the target resp. source language. SUSY/STS applications are possible for the language directions German -> English, English -> German and Russian -> German, some effort has also been made for the development of French -> German and Esperanto -> German. The main application is made in German -> English, where a basic German analysis dictionary of about 150,000 entries, a German compound dictionary of about 140,000 entries, a German -> English dictionary of about 80,000 entries is available.

- As SYSTRAN, SUSY/STS needs an administrative staff to handle the application. Therefore, one concentrates on a translation service concept. Translations are processed only at the computer center of the IAI, Saarbrücken. The clients have to send their data via magnetic tape to the IAI, they get back a magnetic tape including the translations. The main applications are, as mentioned, the translation of German titles coming from different data bases to English. Besides, the bilingual terminology in the different disciplines is updated, the dictionaries are augmented with the special terminology during the processing of the client's data. Clients, besides others, are: The Information Center for Building Construction, where a bilingual data base (ICONDA) is produced; the German Patent Office, where - at the moment - the German Catchword Index is translated to English; the German Institute of Standards, where the titles of German industrial standards are translated to English to be integrated in the corresponding data base; the Information Center for Social Science, where the titles of a bibliographical data base are translated. Up to now, more than 100,000 titles have been translated by this service.

- The System is available under UNIX (and in this sense, portable), the programming language is FORTRAN, with some basic parts written in "C". A version is also available on SIEMENS mainframe computer.

- The quality of the (raw) machine translation (German -> English) is satisfying: about 99 % identification of words in the source language on morphological level; about 90 % correct identifications on the syntactic level. There are no statistics available about the correctness of semantic disambiguation, but development of tools for integrating discipline codes is under progress.

- The service normally integrates the post-editing to reach a high quality of translation. As just mentioned, the client sends its data via magnetic tape to the service center. As a first step, a spelling check is done to identify and correct misspelled words. Then, the unknown words (mainly in the transfer dictionary of the MT system) are identified, the dictionary is then updated (by the human translator). In the next step, the automatic raw translation is done and the text is post-edited (on screen using a PC or terminal) by the human translator. After a controlling phase, the magnetic tape is sent back to the client. It is possible to integrate an indexing system, that means to use the analysis phase of SUSY/STS to produce lemmatized words (instead of word forms) and decompose or derive complex words to get relations between compounds and the single words from which they are derived.

- As SYSTRAN, SUSY/STS merely is batch-oriented. But a special trained user would be able to integrate new words into the different dictionaries: the dictionary maintenance is dialogue-oriented. Post edition can be done seperately on PC or via terminal on the host system (i.e. the mainframe computer, which, at the service center, is a Nixdorf TARGON 35).

- The development cost of the system were about $ 5,000,000 (basic research at the university not included). The adaptation to the STS service environment cost about $ 200,000. The basic cost keeping the system available and running are about $ 150,000

/ year, the service personal (not the translators) included. The service is cost-covering, if about 400,000 titles or sentences can be translated (fee / title: 2 DM = about $ 0.80.

- As a conclusion, one can say that SUSY/STS is a specialized CAT system. It's speed (5,000 running words / hour, CPU time, can be translated) is not comparable with SYSTRAN, and there will not be an augmentation of the language pairs. But it can be shown that, under consideration of the possibilities and by adaptation to special needs, MT systems will play an interesting role in the future.

Under the limits of this article, it is not possible to give similar descriptions of other existing (and available) systems. It also needs a good familiarity in the use and a good knowledge of the conceptual level. The aim was to give some concrete examples to give a first impression of the complexity of this theme. So it seems to be to simple to handle MT from the pure "linguistic" view (in the sense, that the quality of a MT is or is not comparable with the quality of human translation) or - on the other side - to handle MT as a pure technical software tool: Progress can only be made if one considers the limits and the possibilities of MT in concrete application environments.

5. Outlook

If one looks in a distant future, it seems quite clear that professional translation will normally be undertaken by using MT systems. It depends more or less on the availability of the "right" language pairs, the completeness of machine vocabularies, the availability of the "right" technical environment and, last not least, on the cost of the utilization of the system.

In-between (and besides), CAT systems on the level of vocabulary and terminology aid - integrated in or added to text processing systems, will play a substantial part, esp. as an aid during the creation of a text in a foreign language by the author or to understand a written machine readable text by the receiver arriving via telecommunication (Teletex). In the office environment, CAT, mainly with bilingual dictionaries on CD-ROM or hard disk will be used as a variant of style and grammar help.

On the contrary of some assumptions (see, e.g., Hutchins 1986, pp 331 - 334), we don't see any need for a special translator's work station, but there will special functions as translation help be integrated in powerful writer's workstations, so that everybody will make profit of the development in computerized dictionaries and thesaurus.

What will happen with the human professional translator in the environment of technical translations? One can see MT system as a special expert system with the problem, that it will nearly never reach - generally spoken - the full quality of good and specialized human translators. But to develop, to augment and even to maintain such translation expert systems, one needs "language knowledge engineers" which are able to handle and to feed these systems. On behalf of the complexity of natural language and knowledge, there will be cooperation between the system and the translator. And there is a good chance, that if translation of texts will become cheaper and faster, more demand will arise, so that, at the end, human translators will still play a substantial part - under changed conditions.

Literature:

Hutchins, W.J. (1986): Machine translation: past, present, future. Chichester.

Nagao, M. et al (1982): An English-Japanese machine translation system of the titles of scientific and engineering papers. In: COLING 1982, Amsterdam.

Van Slype, G. (1979): Deuxième évaluation du système de traduction automatique SYSTRAN anglais-français de la Commission des Communautés Européennes. Bruxelles.

Van Slype, G. (1982): Conception d'une méthodologie générale d'évaluation de la traduction automatique. Multilingua 1, 221-237.

Zimmermann, H., Kroupa, E., Luckhardt, D. (1987): STS - Das Saarbrücker Übersetzungssystem. Saarbrücken 1987.

Aspects linguistiques et techniques de la traduction automatique

Harald H. Zimmermann
Université de Saarlande
5.5 Informationswissenschaft
D-6600 Saarbrücken 11
RFA

Résumé

Pour permettre de comparer les systèmes de traduction assistée par ordinateur (CAT) et de traduction automatique (MT), les critères essentiels et les caractéristiques des différents concepts sont présentés. Parmi les critères majeurs de différenciation figurent:

- Traduction de textes complets (y compris les titres, les résumés) par opposition à une aide lexicale à la traduction;
- Traduction d'ordre général par opposition à une traduction de textes dans des domaines spécialisés;
- Traduction bilingue par opposition à traduction multilingue;
- Enrichissement des dictionnaires et des systèmes orienté vers l'utilisateur;
- Portabilité.

Parmi les critères essentiels d'évaluation sont:

- Qualité de la traduction brute (aussi appelée informative); Avantages/inconvénients par opposition à la traduction humaine;
- Intégration dans un environnement d'utilisateurs (bases de données, indexation et recherche, traitement de textes, édition électronique);
- Convivialité;
- Coûts du développement et de l'installation;
- Coûts de fonctionnement et effets de réalisation.

Enumération des symboles

CAT	Traduction assistée par ordinateur
CE	Communauté européenne
EUROTRA	Système européen de traduction, projet de recherche de la Communauté européenne
LOGOS	Système commercialement disponible de traduction automatique de LOGOS Corporation;
MINITEL	Outil d'accès du système vidéotex au TELETEL français
MT	Traduction automatique
PC	Ordinateur personnel
SUSY/STS	Système/service de traduction de Saarbrücken
SYSTRAN	Système commercialement disponible de traduction automatique de World Translation Center, Etats-Unis et de Gachot S.A., France
Wang OIS	Systèmes de bureau Wang
WordPerfect	Editeur commercialement disponible

INTRODUCTION

Vu le développement technique de l'industrie de l'information, on doit réaliser que l'ordinateur sera intégré dans presque chaque activité humaine professionnelle. Dans le domaine du traitement de textes, l'intéraction homme-machine, au niveau du PC ou sur la base d'un poste de travail sur terminal au lieu d'une machine à écrire, sera la norme dans un proche avenir, à savoir d'ici à la fin des années 80.

D'autre part, le problème de la traduction automatique de texte naturel en général — sans même parler de la compréhension de la parole et de la traduction de la parole — ne sera pas résolu dans le sens de la HQFT, (traduction entièrement automatique de haute qualité), en raison de la complexité des langages naturels. Il y a donc des limites (plus ou moins dans l'approche scientifique) et des possibilités d'intégration des outils techniques dans le cours de la traduction (humaine) ou même de la compréhension du texte par la machine.

Le but de cette contribution est d'explorer les possibilités de l'utilisation de l'ordinateur dans le domaine de la traduction automatique et de la traduction assistée par ordinateur. Même s'il serait intéressant de traiter également les principaux problèmes de traduisibilité des textes (et de la compréhension), nous nous concentrerons sur la traduction technique, c'est-à-dire la traduction de textes techniques ou communs.

Aspects systématiques

Même s'il pourrait être intéressant de savoir comment la traduction automatique est réalisée, les aspects linguistiques (en particulier les modèles de la grammaire et des stratégies possibles ou employées) ont un rôle secondaire. Aussi on peut

supposer qu'il y a une sorte de "boîte noire", où un texte ou des mots dans un langage naturel sont insérés et, avec ou sans intéraction humaine, une ou plusieurs possibilités de traductions. La traduction pourrait être bonne, utile ou mauvaise selon les avis des utilisateurs.

En se concentrant sur l'utilisation et l'utilité de la traduction automatique ou de la traduction assistée par ordinateur, on doit — en principe — faire la distinction entre deux groupes d'utilisateurs principaux:

(1) l'utilisateur dit final, c'est-à-dire un spécialiste qui essaie de savoir de quoi parle un article écrit en langage naturel plus ou moins inconnu de lui; une personne écrivant une lettre à un ami en langue étrangère... et

(2) le professionnel, en particulier un traducteur humain qui essaye d'utiliser la machine comme un outil dans l'accomplissement de son travail.

Vu sous cet angle, on peut laisser de côté les projets et la recherche fondamentale sur les principes de la traduction automatique / compréhension des langues et s'intéresser plus particulièrement aux outils, systèmes pratiques.

Comme point de départ, il est important de faire la distinction entre deux stratégies principales: traduction automatique (MT) et traduction assistée par ordinateur (CAT).

— Un système s'appellera un système de traduction automatique (MT) seulement si le processus de traduction — en partant du texte source lisible sur ordinateur — est accompli sans aucune intéraction humaine pour atteindre une qualité de texte cible au moins "assez bonne" pour l'information.

— Un système s'appellera un système de traduction assistée par ordinateur (CAT) si une intéraction humaine est nécessaire ou prévue pour atteindre le but d'une "bonne" traduction d'un texte source (lisible sur ordinateur ou pas).

Il est bien évident qu'un système de traduction automatique peut venir en complément d'un système de traduction assistée par ordinateur: un texte peut être amélioré avant de débuter la traduction automatique de façon à obtenir une meilleure traduction automatique ("pré-édition"), et/ou un texte traduit par ordinateur peut être post-édité par un traducteur humain pour obtenir une meilleure qualité.

Il existe sur le marché de nombreux systèmes de traduction automatique et de traduction assistée par ordinateur qui prétendent être les véritables outils pratiques, et le choix est difficile à faire sans avoir de critères précis pour en décider.

Puisque la quantité de données dans les dictionnaires automatisés — sur la longue période de développement des systèmes MT et CAT — est le composant décisif, la mise à jour du dictionnaire (électronique) joue un rôle substantiel dans les deux options suivantes:

— Un concept peut être — en particulier si le système a encore une grande capacité d'entrées dans le dictionnaire — que l'utilisateur, par exemple le traducteur ne devrait pas participer à un tel processus (voir, par exemple le concept Systran), de sorte qu'il y ait des spécialistes requis du côté de l'administration de système pour améliorer la base de données du dictionnaire;

— L'autre manière est de laisser le soin à l'utilisateur professionnel de compléter le dictionnaire du système ou y ajouter un vocabulaire spécial (voir, par exemple, le concept Logo).

Il y a d'autres aspects du système qui jouent un rôle important dans le processus de décision: la disponibilité des couples de langues, la possibilité d'utiliser ou de traiter des types de texte spéciaux (par exemple, minutes, lettres,...).

Critères d'évaluation

Qualité

Même si on peut (ou doit) normalement traiter le système comme une "boîte noire", il existe des différences dans la qualité des résultats "bruts" de traduction automatique (sortie). Il n'est pas très facile de donner une mesure précise, mais il y a certains critères importants à noter (voir, pour les détails, en particulier le concept de Van SLYPE (1982) et les descriptions des deux évaluations SYSTRAN, Van SLYPE (1979).

Les critères principaux à considérer sont:

(1) fiabilité et fidélité, ce qui signifie: à quel degré (mauvais, assez bon, bon) la teneur/signification du texte original est conservée.
(2) compréhensibilité et intellegibilité, ce qui signifie: à quel degré l'utilisateur final peut lire et comprendre le texte traduit.

Ceci paraît assez simple, mais les problèmes se produisent dans le détail; Ainsi, au niveau syntaxique et stylistique, un système produira normalement des traductions qui sont pires que les traductions humaines, tandis que au niveau lexical, particulièrement dans l'identification du(des) bon(s) terme(s) technique(s), la traduction machine peut même être plus précise et consistante qu'une traduction humaine.

Environnement des possibilités d'application.

Un système de traduction automatique ou de traduction assistée par ordinateur doit être considéré dans l'environnement (technique) des possibilités d'application. Ainsi, normalement une décision est non seulement orientée vers la qualité, mais également basée sur les possibilités d'intégration dans un système de traitement de texte. Vu sous cet angle, les éléments suivants doivent être considérés:

Intégration dans des bases de données (bibliographiques ou textuelles).

Entretemps, les bases de données sont — techniquement parlant — mondialement accessibles par l'intermédiaire de réseaux de commutation et même de satellites de communication. Ainsi, réduire les barrières linguistiques, par exemple entre l'anglais et le Japonais, mais particulièrement sur le marché multilingue européen, devient un vif souhait. Des efforts expérimentaux sont consacrés au Japon pour intégrer les systèmes de traduction automatique dans un tel processus de l'information: par exemple la base de données INSPEC qui est originellement anglaise peut être interrogée avec les mots clés japonais. Les mots clés et plus tard les titres (anglais) sont automatiquement traduits au cours du dialogue en japonais (Nagao et autres 1982). D'une manière semblable, un système de traduction automatique de l'allemand en anglais est utilisé en traitement par lots avec post-édition pour traduire les titres des bases de données allemandes (Zimmermann et autres 1987).

Il est bien évident que la traduction de titres et de résumés pourrait également être faite par des traducteurs humains. Mais il y a des arguments en faveur des systèmes de traduction automatique et de traduction assistée par ordinateur: le texte à traduire est lisible par ordinateur, il s'agit donc d'une base idéale pour utiliser un ordinateur, les domaines ou les secteurs vers lesquels la rubrique/le texte est orienté sont normalement "physiquement" marqués, de sorte que la classification ou même les fonctions de thésaurus puissent être employées, particulièrement pour le transfert lexical (désambiguisation), le vocabulaire technique demande à être très précis, de sorte que l'ordinateur aide en étant homogène.

Indexage automatique.

L'indexage (complet) d'un texte peut être une application secondaire valable de l'utilisation des systèmes de traduction. En effet, pour le transfert lexical, on doit faire dériver les formes de mots des formes de base, on doit identifier les mots composés et les mots complexes en tant que tel, les catégories de mots, voire les relations entre les termes aux fins de désambiguisation. Ainsi peut-on envisager qu'un système de traduction soit utilisé de façon définitive ou intermédiaire pour l'archivage de documents et la recherche documentaire.

Traitement de texte.

Il n'y a aucun doute que le traitement de texte joue un rôle important dans chaque environnement de traduction. Même les traducteurs (humains) indépendants utiliseront de plus en plus un système de traitement de texte (sur PC), et l'on tend, petit à petit, à intégrer, d'une manière ou d'une autre, les glossaires ou les listes (personnels) et de les rendre accessibles par l'intermédiaire de "fenêtres" — au lieu d'utiliser les fichiers manuels. Il est bien évident que d'autres fonctions, par exemple, les contrôleurs orthographiques, seront de plus en plus intégrés dans un tel processus.

En conséquence, le "texte source" est assimilable par ordinateur, mais on doit adapter les systèmes de traduction aux différents traitements de texte (par exemple Wang OIS est combiné aux systèmes LOGOS et SYSTRAN, WordPerfect à SYSTRAN). Si de tels outils sont disponibles, la post-édition de la traduction brute pourra être réalisée par un réviseur confirmé.

Un des problèmes qui se pose dans cet environnement est la combinaison de ces outils avec les systèmess locaux de MT (par exemple, LOGOS) ou la connexion à un serveur de traduction (voir l'application de SYSTRAN au sein de la Commission européenne ou dans le réseau Minitel en France pour obtenir des traductions automatiques.

Il ne fait aucun doute que l'utilisation des systèmes de traduction progressera principalement en parallèle avec le traitement de textes et l'accès en direct. La question est, pour l'instant, de savoir si les outils existants sont assez puissants (en qualité) pour être acceptés par l'utilisateur. En particulier, les résultats de l'expérience de TAO menée par GACHOT S.A. sur Minitel et via PC (avec SYSTRAN) constitueront une appréciation importante à cet égard.

Edition électronique

De plus en plus, les textes à traduire (en particulier, les textes techniques comme les instructions de réparation, les manuels) sont entièrement préparés — y compris les chiffres, les plans, les tableaux, les images — par l'intermédiaire de l'édition électronique, que l'on connait aussi sous le nom de micro-édition. Les societés qui utilisent les services de traducteurs, internes ou externes, ne veulent pas remettre en forme (ou composer) une traduction même dans beaucoup de langues cible.

Aussi un grand effort doit-il être consacré (et ce, de la part du fabricant de systèmes d'édition électronique et du producteur de logiciels de traduction) pour intégrer les aides à la traduction sans supprimer ou violer la structure du document. Naturellement, il existe des problèmes d'ajustement de ligne, de paragraphe ou de page (dus à la différence de longueur entre textes source et cible), et, le réarrangement des expressions/des mots du fait des différences syntaxiques entre langues, soulève nombre de difficultés quant à l'intégration correcte des descripteurs typographiques (caractères gras, soulignement, etc.) dans la langue cible. Si l'on ne souhaite pas confier ces éléments "techniques", qui font partie de la traduction, au traducteur/réviseur, il convient d'intégrer aux systèmes de micro-édition des moyens plus performants de description de la forme du texte. Voilà le défi que les systèmes de TAO existants et futurs doivent relever.

Convivialité.

Pour le poste de travail du traducteur, la convivialité joue un rôle important à chaque niveau des systèmes de traduction. Pas un traducteur humain ne souhaiterait utiliser un système de traduction et devoir jouer le rôle d'esclave qui consisterait à corriger jour après jour, les mêmes erreurs "grossières" de traduction du système. Le danger existe à l'heure actuelle, parce que les systèmes existants ne sont pas suffisamment flexibles et adaptés.

Les développements futurs des systèmes de traduction doivent donc se concentrer sur les activités qui donnent à l'utilisateur davantage de possibilités de rétroaction. Le "concept du fichier personnel" qui est parfois employé dans les systèmes de base de données, où un utilisateur peut choisir et créer une "partie personnelle" dans une base de données, pourrait être un exemple: au moins au niveau du dictionnaire, l'utilisateur devrait disposer de fonctions pour réaliser — sur la base des données existantes — son ou ses propres dictionnaires (physiquement ou logiquement). A et égard, ausssi bien les fabricants de systèmes de base de données et les utilisateurs, peuvent tirer profit d'un tel concept; le vocabulaire spécialisé du système sera amélioré et l'utilisateur aura une influence non négligeable (mais aussi responsable) dans le choix des traductions.

Ce qui est vrai pour la partie lexicale, devrait également être applicable aux éléments structurels (par exemple. pour agir sur la longueur des phrases, ou certains composants stylistiques, etc...). Les systèmes existants devraient offrir davantage de flexibilité et les systèmes futurs devraient prendre en considérations de tels éléments dès le commencement.

Rentabilité.

Commercialement parlant, la décision d'utiliser un système de traduction est basée sur un calcul de rentabilité. Dans ce cas, ce n'est pas uniquement un problème financier pur, parce que le "temps c'est de l'argent et qu'une traduction obtenue sans tarder peut avoir un prix plus élevé.

Mais en fin de compte, la décision est prise sur une base économique, en tenant compte des effets sociaux ou humains.

Partant du fait que les renseignements précis sont rares sur le coût de développement des systèmes de TAO, pour autant que l'auteur le sache, nous devons étudier la question de la rentabilité au niveau de l'utilisateur.

Il semble, à l'heure actuelle, que la quantité de traductions (mesurées en pages/jour) qui sont produites par un traducteur humain par intéraction et/ou post-édition peut être sensiblement augmentée. En supposant que, par exemple, la maintenance informatique et l'application technique d'un système comme SYSTRAN coûteront $300,000/ année, et que 300,000 pages peuvent être techniquement traduites par le système, le coût de la traduction pure — non compris la mise à jour des dictionnaires et la préparation ou post-édition de la traduction brute — est négligeable ($ 1/page). Le coût du processus complet (traduction avec intéraction homme/machine) diffère selon la qualité souhaitée. Pour une traduction d'assez bonne qualité — ce qui signifie une traduction informative par exemple, pour les documents de travail, une traduction rapide peut être faite, de sorte qu'un traducteur humain produise environ 20 pages/jour (au lieu de 4-6 pages sans TAO).

Pour atteindre un haut niveau de qualité comparable aux traductions humaines professionnelles, on doit consacrer davantage de temps à la révision ou post-édition. Mais il semble que le point d'équilibre soit atteint dans ce sens que si le vocabulaire d'un système de traduction est adapté au domaine de l'utilisateur, le coût de traduction est inférieur au coût de traduction humaine pure, même en considérant que l'utilisation des systèmes de traitement de textes dans le processus de traduction fait gagner environ 20 % du temps.

Problèmes linguistiques et stratégiques de base.

L'aspect morphologique (à savoir les problèmes d'inflexion, de dérivation, et de décomposition) des systèmes de traduction ont été résolus avec succès, au moins à des fins pratiques d'environnement d'application, même si la traduction correcte des mots dérivés et décomposés n'est pas toujours atteinte au moyen de règles automatisées. Ce n'est pas le cas pour des solutions aux niveaux syntaxique et sémantique. Même si nous supposons qu'un problème, par exemple, la désambiguisation des homographies syntaxiques (comme RAINS dans IT RAINS ou THE RAINS) peut être résolue grâce à un système d'analyse strict et entièrement formalisé, la complexité de la structure du langage naturel mène à un dépassement du temps d'ordinateur si on essaye d'intégrer ou de manipuler chaque structure ou occurence (partielle). Ainsi, en réalité, les systèmes commercialement disponibles essayent de raccourcir le processus d'identification (ou de disambiguisation) par le truchement de règles déterministes particulières ou de probabilités. En conséquence, les systèmes fonctionnent 10000 ou 1000 fois plus rapidement qu'un système entièrement basé sur la linguistique, mais leurs résultats ne peuvent pas atteindre la même qualité.

Aujourd'hui, le temps d'ordinateur ne joue plus le même rôle qu'il y a quelques années, mais le temps d'ordinateur de traduction automatique jusqu'à maintenant n'est pas entièrement négligeable. Ceci est vrai aussi pour la solution des problèmes d'homonynie, à savoir dans le domaine sémantique. D'une part, il y a des limites dans le traitement de la structure des textes par opposition à la structure de la phrase. Dans la plupart des cas, la base de connaissances d'un système de traduction est l'environnement de la phrase, ce qui signifie que les informations ou les solutions des phrases précédentes ne sont pas prises en compte et que presque aucune donnée n'est connue au niveau des textes. Ceci soulève beaucoup de problèmes, en particulier dans le domaine de la référence pronominale, mais également dans l'insertion d'article et la désambiguisation des homonymes. Dans les systèmes existants, la résolution du problème général des ambiguïtés sémantiques passe par l'introduction de codes sémantiques (qui — sur un plan général) — joue également un certain rôle dans la désambiguisation des des structures syntaxiques), particulièrement les marqueurs par domaine ou par discipline, qui sont utilisés pour choisir le "bon" mot (ou séquence de mot) selon les paramètres spécifiques données par l'utilisateur. Ils essayent également de résoudre ce problème

par une consultation du dictionnaire afin d'identifier les séquences de mots ou même les citations (qui doivent normalement être lexicalisées parce qu'il n'existe aucune solution algorithmique), voir l'exemple en allemand de ES REGNET BINDFADEN, qui doit être traduit par IT RAINS CATS AND DOGS.

Depuis Chomsky, l'approche structurale-sémantique systématique à l'analyse de la langue, à la "compréhension" et à la traduction a progressé. Ainsi, au niveau de la recherche, on dispose maintenant de plusieurs types de grammaire formalisés et de systèmes d'analyse disponibles. Les recherches en traduction automatique continuent, que ce soit au Japon (voir le MU-SYSTEM), ou en Europe (à savoir les efforts consacrés par la Communauté européenne et les Etats membres pour le système européen de traduction EUROTRA). Mais le développement linguistique informatique n'est pas suffisant: les linguistes, les informaticiens, les documentalistes et les utilisateurs doivent coopérer dans les projets à grande échelle pour viser le but d'une utilisation pratique.

Exemples

Pour donner quelques notions sur l'état de la question des systèmes dits productifs (pour ne pas dire commercial), et pour montrer également comment les critères mentionnés peuvent être appliqués, on décrira deux systèmes, les systèmes SYSTRAN et SUSY/STS.

SYSTRAN

SYSTRAN (droits commerciaux à Gachot S.A., France) — dans sa nouvelle version 3.7 — a les caractéristiques suivantes:

— Traduction de texte intégral. Même si la structure de la phrase n'est pas correcte, si des mots sont mal orthographiésu ou s'ils ne figurent pas dans le dictionnaire du système, une traduction est faite.

— La vitesse de traduction (selon la capacité d'ordinateur) atteint jusqu'à 35O.000 mots/heure. C'est donc le système le plus rapide disponible sur le marché.

— Le système fonctionne par paires de langue. Les traductions sont effectuées dans les paires de langues: anglais — français, anglais — italien, français — anglais, russe — anglais (USAF), anglais — Japonais (SYSTRAN JAPON), anglais — arabe. Les paires de langues actuellement développées sont, entre autres, anglais — allemand, français — allemand, allemand — anglais et allemand — français. La qualité dépend, d'une part, de la disponibilité des dictionnaires spécifiques. Un effort considérable a été consacré à la Communauté européenne pour développer les dictionnaires SYSTRAN. Pour la traduction anglais-français, le dictionnaire contient maintenant plus de 150.000 entrées. La même qualité n'est pas atteinte, par exemple, pour la traduction allemand-français, qui en est à sa phase de démarrage, même au niveau de l'analyse de la langue.

— L'utilisation de SYSTRAN requiert des spécialistes techniques (et des gestionnaires). Ainsi seules des sociétés qui disposent d'un personnel spécialisé (telles que la CEE ou USAF) peuvent utiliser SYSTRAN sur leur propre ordinateur (si l'ordinateur est une unité centrale IBM ou compatible IBM). Mais il y a une alternative intéressante: utiliser le système par l'intermédiaire de réseaux de télécommunication, par exemple réseau de commutation ou — dans une optique très "futuriste" — par vidéotex. En France, une application vidéotex (appelée TELETEL) en utilisant un téléphone plus un écran (MINITEL) est déjà opérationnelle (et est même employé par les élèves).

— L'enrichissement du dictionnaire. Jusqu'ici, les dictionnaires SYSTRAN sont normalement faits par les spécialistes du système. Le problème principal n'est pas le codage lui-même (qui est très complexe, mais pourrait et sera résolu grâce à des interfaces conviviales), mais la cohérence de la base de données du dictionnaire. Les dictionnaires contiennent des codes de domaine, qui demandent à être améliorés de façon à mieux correspondre aux attentes de l'utilisateur.

— La portabilité de SYSTRAN est presque impossible à réaliser: SYSTRAN est écrit en IBM-ASSEMBLEUR, même si les règles linguistiques sont normalement écrites dans des macros spéciales. Le système lui-même fonctionne — comme mentionné ci-dessus sur unité centrale (IBM ou SIEMENS ou AMDAHL) ou sur des ordinateurs de dimension semblable. Une révision de logiciel (peut-être sur UNIX) est projetée.

— La qualité de traduction (brute) diffère selon les couples de langues et la capacité du dictionnaire: Si on se concentre sur le français-anglais, les pourcentages suivants peuvent s'appliquer: Identification morphologique: environ 100 %; structures syntaxiques: environ 90 %; désambiguisation sémantique: entre 80 et 90 %, selon l'intégration du dictionnaire des LS (limited semantic).

— SYSTRAN peut être combiné à plusieurs environnements de traitement de texte. On peut utiliser un Wang OIS (employé à la CEE), mais aussi utiliser un PC (IBM compatible) avec un WordPerfect. De même, on peut saisir les textes en lecture optique et créer des fichiers-texte avec WordPerfect, envoyer les données au serveur SYSTRAN (par exemple à Gachot S.A., Paris) et récupérer le texte traduit en différé. Enfin, un outil-logiciel permet de post-éditer le texte sur écran divisé (visualisation simultanée du texte source et cible).

L'utilisation du système de traduction elle-même se fait en traitement par lots. Pendant le processus de traduction, il n'y a aucune possibilité d'intervention (en dépit du fait que le gestionnaire du système a la possibilité d'introduire des traductions pour les mots inconnus). D'autre part, l'utilisateur utilise le système comme une boîte noire, il n'a besoin d'aucune connaissance du système.

Il n'existe aucune information disponible sur le coût de développement du système. On peut estimer qu'il se situe entre $20.000.000 et $50.000.000. Rien que pour la CEE, le coût d'utilisation est de l'ordre de $4.000.000 à $6.000.000. Le coût d'exploitation de la version SYSTRAN (sans les royalties) avoisine — à mon avis — les $300.000/an. Aussi si l'on prend l'hypothèse de 300.000 pages traduites par an (par exemple, à la CEE), le coût est inférieur à $1/page (à l'exception du traitement de texte, de la post-édition et du temps d'ordinateur). Le coût de traduction sur MINITEL (vidéotex) est d'environ $0,10 par "fenêtre" (jusqu'à 10 lignes) et la traduction (brute) automatique s'affiche normalement après une attente de 45 à 60 secondes.

Ceci conduit à la conclusion que SYSTRAN est conçu pour l'utilisateur, à savoir que son marché est celui de la "société de l'information où il n'est pas indispensable que la qualité atteigne 100 %, mais qu'un utilisateur qui ne connaît pas ou pas suffisamment la langue ait une bonne compréhension du texte traduit.

SUSY/STS.

SUSY/STS, développé dans le cadre d'un grand projet de recherche en linguistique computationnelle à l'université de Saarbrück (RFA), est maintenant employé comme système de production dans le domaine de la traduction de bases de données. Ce service est fait par l'Institut de la Recherche de l'Information Appliquée (IAI) à Saarbrück. Les principales caractéristiques sont:

— La traduction de textes intégraux (au niveau de la phrase) est possible. Dans l'environnement d'application, on se concentre sur la traduction du titre (et du résumé).

— Comme SYSTRAN, SUSY/STS est un gros système de traduction de portée générale: un texte est traduit même s'il existe des mots inconnus. Dans un tel cas, le mot non connu est restitué dans le texte cible.

— Dans une certaine mesure, SUSY/STS est un système de traduction multilingue. Seul la phase de transfert (entre deux langues) est bilingue, tandis que les phases d'analyse et de synthèse sont effectuées indépendamment de la langue source et cible. Les applications SUSY/STS sont possibles pour les langues Allemand — anglais, anglais — allemand et russe — allemand, et des efforts sont consacrés au développement du français — allemand et espéranto — allemand. L'application essentielle est faite en allemand — anglais, avec un dictionnaire de base d'analyse de l'allemand d'environ 150,000 entrées, un dictionnaire d'expressions d'environ 140.000 entrées, et un dictionnaire allemand — anglais d'environ 80,000 entrées.

— Comme SYSTRAN, SUSY/STS exige un personnel spécialisé pour faire fonctionner l'application. Par conséquent, on se concentre sur un concept de service de traduction. Les traductions sont effectuées uniquement au centre de calcul de l'IAI, à Saarbrück. Les clients doivent envoyer leurs données sur bandes magnétiques à l'IAI, ils obtiennent en retour une bande magnétique comprenant les traductions. Les applications principales sont, comme mentionné, la traduction en anglais de titres allemands venant des différentes bases de données. En outre, la terminologie bilingue dans les différents domaines est mise à jour, les dictionnaires sont enrichis avec la terminologie spécifique pendant le traitement des données du client. Les clients, entre autres, sont: Le Centre d'Information pour la Construction du Bâtiment, où une base de données bilingue (ICONDA) existe; l'Office des Brevets allemand où — à l'heure actuelle — l'Index des mots-clés en allemand est traduit en anglais; l'Institut allemand des Normes, où les titres des normes industrielles allemandes sont traduits en anglais pour être intégrés dans la base de données correspondante; le Centre d'Information pour les Sciences Sociales, où les titres d'une base de données bibliographiques sont traduits. Jusqu'ici, plus de 100.000 titres ont été traduits par ce service.

— Le système fonctionne sous UNIX (et est, dans ce sens, portable). Le langage de programmation est FORTRAN, avec certaines parties de base écrites en "C". Une version est également disponible sur l'ordinateur central de SIEMENS.

— La qualité de la traduction automatique (brute) (allemand — anglais) est satisfaisante; environ 99 % d'identification des mots dans la langue source au niveau morphologique; environ 90 % d'identifications correctes au niveau syntaxique. Il n'y a aucune statistique disponible sur l'exactitude de la désambiguisation sémantique, toutefois la mise au point d'outils permettant d'intégrer les codes de domaines est en cours.

— Le service assure normalement la post-édition pour atteindre un haut niveau de qualité de traduction. Comme précisé ci-dessus, le client transmet ses données sur bande magnétique au centre de calcul. Dans un premier temps, on procède à un contrôle orthographique pour corriger les mots mal orthographiés. Ensuite, les mots inconnus (essentiellement dans le dictionnaire de transfert du système de traduction) sont identifiés, le dictionnaire est alors mis à jour (par le traducteur humain). Le texte est ensuite traduit automatiquement et post-édité (sur écran à l'aide d'un PC ou d'un terminal) par le traducteur humain. Après une phase de contrôle, la bande magéntique est retournée au client. Il est possible d'intégrer un système d'indexage, c'est-à-dire de partir de la phase d'analyse de SUSY/STS pour engendrer des mots lemmatisés (au lieu de formes de mots) et décomposer ou dériver des mots complexes pour obtenir des relations entre les mots composés et les mots simples desquels elles sont dérivées.

— Comme SYSTRAN, SUSY/STS fonctionne simplement en traitement par lots. Mais un utilisateur spécialement formé pourrait intégrer des mots nouveaux dans les différents dictionnaires: l'enrichissement de dictionnaire se fait en mode conversationnel. La post-édition peut être faite séparément sur PC ou par un terminal sur l'ordinateur central (Nixdorf TARGON 35 au Centre de calcul).

— Le coût de développement du système est de l'ordre de $5,000,000 (non inclus la recherche de base à l'université). L'adaptation à l'environnement de service STS se chiffre à environ $200,000. Le coût de base de fonctionnement de

l'ordinateur est de $150,000/an, ce coût incluant le personnel (pas les traducteurs). Le service atteint son seuil de rentabilité à partir du moment où environ 400.000 titres ou phrases ont été traduits, (prix d'un titre: 2 DM = environ $0,80).

— Comme conclusion, on peut dire que SUSY/STS est un système de traduction spécialisé. Sa vitesse (traduction de 5.000 mots/heure, temps CPU) n'est pas comparable à SYSTRAN, et il n'y aura pas d'augmentation des couples de langues. Toutefois, on peut présumer que, en prenant en compte les possibilités et l'adaptation aux besoins spécifiques, les systèmes de traduction joueront un rôle intéressant à l'avenir.

Dans les limites de cet article, il n'est pas possible de donner une description précise des autres systèmes de traductions existants (et disponibles). Ceci exige aussi d'être au fait de l'utilisation des systèmes et d'avoir une bonne connaissance au niveau conceptuel. Le but de cet article était d'exposer certains exemples concrets afin de donner une première impression de la complexité de ce thème. Ainsi il peut paraître aisé d'utiliser un système de traduction du point de vue "linguistique" pur (dans le sens que la qualité de traduction machine n'est pas comparable à la qualité d'une traduction humaine) ou — d'autre part, d'utiliser un système de traduction comme un outil-logiciel technique pur. Des progrès ne seront accomplis que dans la mesure où l'on tient compte des limites et des possibilités des systèmes de traduction dans des environnements d'application concrets.

5. Perspectives.

Si l'on se tourne vers un avenir lointain, il semble assez clair que la traduction professionnelle passera par des systèmes de traduction. Ceci dépendra plus ou moins de la disponibilité des "bons" couples de langues, de l'état complet des dictionnaires, de la mise en place du "bon" environnement technique et, argument non des moindres, du coût d'utilisation du système.

En attendant, les systèmes de traduction, en tant que systèmes d'aide terminologique à la traduction — intégrés ou ajoutés aux traitements de texte, joueront un rôle important, particulièrement comme aide à la création d'un texte dans une langue étrangère par le rédacteur ou comme aide à la compréhension d'un texte lisible par ordinateur par le destinataire par télécommunication (Teletex). Dans l'environnement bureautique, les systèmes de traduction assistée par ordinateur, principalement grâce aux dictionnaires bilingues sur CD-ROM ou sur disque dur seront utilisés comme une variante d'aide de style et de grammaire.

Contrairement à certain s hypothèses (voir Hutchins 1986, p.331-334), nous ne voyons pas la nécessité d'un poste de travail traducteur, mais plutôt certaines fonctions spécialisées telles que l'aide à la traduction intégrée dans un poste de travail du rédacteur, de sorte que chacun tire profit de la mise au point de dictionnaires et de thésaurus automatisés.

Quel sera le rôle du traducteur humain dans l'environnement de la traduction technique ? On peut voir un système de traduction comme un système spécialisé qui toutefois, n'atteindra sans doute jamais — d'une manière générale — le haut niveau de qualité des traducteurs spécialisés. Mais pour mettre au point, améliorer et même maintenir de tels systèmes de traduction, on a besoin d'ingénieurs spécialisés en langues capables de faire fonctionner et d'enrichir ces systèmes. La complexité du langage naturel et les connaissances requises exigent une cohésion entre le système et le traducteur. Et il y a fort à penser que si la traduction de textes devient meilleur marché et plus rapide, la demande augmentera, de sorte que, en fin de compte, les traducteurs continueront à jouer un rôle substantiel — dans des conditions différentes.

Bibliographie:

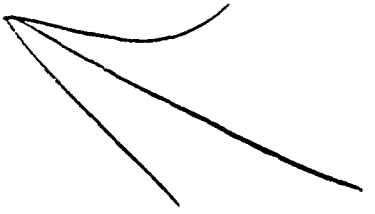
Hutchins, W.J. (1986): Machine translation: past, present, future. Chichester

Nagao, M et autres (1982): An English-Japanese machine translation system of the titles of scientific and engineering papers. In: COLING 1982, Amsterdam.

Van Slype, G. (1979): Deuxième évaluation du système de traduction automatique SYSTRAN anglais-français de la Commission des Communautés Européennes. Bruxelles.

Van Slype, G. (1982): Conception d'une méthodologie générale d'évaluation de la traduction automatique, Multilingua 1, 221-237.

Zimmermann, H., Kroupa, E., Luckhardt, D. (1987): STS — Das Saarbrücker Ubersetzungs-system. Saarbrücken 1987



Information Retrieval Systems Evolve - Advances for Easier and More Successful Use

Gladys Cotter
Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22304-6145

ABSTRACT

This paper describes the advances which are being made in information retrieval systems to assist end-users and information specialists to overcome the critical barriers which make it difficult for them to exploit the power of these systems. Technology which is being applied to allow users to interact with information retrieval systems with greater ease and more successful results is identified. To illustrate this process, the efforts of the U.S. Defense Technical Information Center to develop and implement an integrated, functional scientific and technical information network are described. This network was purposefully designed to incorporate both end-users and their information management intermediaries in a complementary manner making them resolute partners in the work and its rewards. The uses of technology modules -- artificial intelligence, expert systems, gateways, user-friendly interfaces — to overcome user barriers are described in the paper.

INTRODUCTION

The media hails the arrival of the information age, and we are told that information is the power of the future and that we are in the midst of an Information Revolution. From the vantage point of the information industry, the view is much more evolutionary in nature. Information has always equated to power, and we are exposed on a daily basis to the gap between media-induced expectations and reality.

The fact that information is a growth industry is indisputable. The number of online databases has increased from 301 in 1976 (Ref. 1) to more than 2800 in 1986 (Ref. 2). In the 1970s, several online services supported these databases. Currently, there are over 400 online services (Ref. 3). Advances in data storage and retrieval capabilities have made it possible to store vast volumes of information online and to retrieve that information in seconds.

The proliferation of desktop computers has led to the vision of a "tactical desktop." Through the tactical desktop, all the information the desk occupant requires would be available by hitting a key, squeezing a mouse, pointing a finger, or issuing a voice command. This, in turn, has led to speculation that information retrieval experts soon will be without purpose, technologically obsolete. Examination of the facts reveals quite the opposite is true. New technology will strengthen the partnership between end-users and intermediaries. Through a combination of technology, imagination, and hard work we will step forward towards the ultimate information system.

BACKGROUND

The information retrieval systems environment consists of end-users, information specialists, database producers, and information systems vendors, using personal computers often linked into office automation systems. These information specialists, end-users, producers and vendors have been working together for years. Recent technological advances in the information systems field, the prevalence of personal computers in the office, and the introduction of office automation systems are changing the roles of, and the interaction among, the players. The culture of the information retrieval market is evolving, but the goal of the culture remains the same -- provide needed information in the most efficient and effective manner.

End-users are the individuals who need information. End-users run the gamut from student to chief executive officer, manager to technical expert. Their reasons for wanting information are as varied as the type of information they require and the format in which they find it suitable. Some end-users want a quick but accurate answer or fact while others merely want clues to aid them in their own discovery process.

Information specialists are intermediaries. Their expertise lies in identifying, retrieving, and delivering needed information to their client population -- the end-users. Information specialists have many tools at their disposal, the most powerful of which are the online retrieval systems. Through these systems, they can manipulate massive amounts of data, sifting through the volume and mining the relevant information with deft precision.

Database producers collect information and are responsible for the content of a database. Database producers may collect information of a certain type, such as all technical reports produced by the United States Department of Defense, or may collect information in a certain subject area, such as geology. Some database producers perform indexing and abstracting services while others provide full-text information. Database producers make their information available through information system vendors. In some cases, the producer and the vendor may be one and the same.

Information system vendors provide the software and hardware to allow users to access databases online. Vendors control the quality and capabilities of the options available for retrieving the information. Database producers also control, to a certain extent, the search options which can be exercised in conjunction with their files. For example, a title search option is useless against a file which contains no titles.

Several years ago, the natural order of the information field was that database producers and vendors marketed their services and products primarily to the information intermediary, who was familiar with the content of the database and the retrieval mechanics provided by the system vendors. The information intermediary, in turn, marketed retrieval from information systems as one of the services available to end-users.

The market at that time was characterized by:

o Homogeneous system user characteristics. Users were information specialists who would devote the time to learn how to use online systems proficiently. Expertise in this area was a professional prerequisite.

o Complex systems. An online session required many commands, from logon to retrieval to printing.

o Unforgiving systems. Omission of a carriage return could be a fatal error.

o Extensive training requirements. There were multi-volume instruction manuals, multi-day introductory courses, and numerous advanced courses.

o Lack of standardization. If all systems are complex, a user who has mastered one will easily become "brand-loyal."

o Rapidly changing state-of-the-art. Information professionals struggled to stay abreast of the new databases and new information systems continually being introduced.

o A unique professional, the information specialist. Terminals and microcomputers were essential tools for their profession.

The essence of the market philosophy was that the information was available for those who had the time to dedicate or the need to develop the skills required to unlock the information banks. At the time, that was a sound philosophy.

Revolution Turned Evolution

The introduction of personal computers into the office (end-user) environment, combined with the developing awareness in end-users that the information available online could provide a competitive edge, played a major role in altering the above-stated philosophy. (The perception that the personal computer would become as commonplace as the telephone in the office and the home environment was probably more important than the reality.)

It did not take long for database producers and vendors to have visions of end-users armed with personal computers performing their own searching. Think how many hundreds of end-users exist for each information specialist. And, we all know that the amount of information end-users ask for is a function of the distance from their desk to the location at which the question must be posed. Reducing the distance between chair and answer to zero could have a tremendous effect on the end-users information needs.

The idea of doing their own searching appealed to many end-users. They were aware of the quantity of information available online to help them meet their goals. They could have all the information they ever wanted at their fingertips. The Information Age had arrived.

End-users, database producers, and vendors rushed to embrace each other. "The End Users are Coming!" became a familiar cry throughout the information community. Information intermediaries were told their jobs were at risk as end-users learned to use remote terminals to unlock information sources.

Yet, for all of this fanfare, the predicted multitudes of end-users have not arrived yet, and direct user accessing of online information remains an almost arcane art. The encounter between the end-users, in their new role, and the information world yielded an important lesson:

Lesson One: As a general rule, end-users, unlike information intermediaries, will not devote extensive time learning to use a system.

End-users found that walking to the library was a lot faster than learning to use the retrieval systems.

Vendors then set out on a quest to find the perfect user interface. To attract end-users, the vendors would have to focus on making these interfaces "user-seductive." Several systems deemed to be "user friendly" were put on the market. More lessons were learned from this experience:

Lesson Two: Lesson One is the only general rule that applies to end-users across-the-board.

Lesson Three: End-users are a diverse group with many diverse needs. They require different types of data, different products and services, different user interfaces, and different marketing strategies.

Lesson Four: User-friendliness is an attribute which is bestowed by the user and not an innate state of being which is created through system design.

The interfaces that failed early suffered from the fact that the market was in its infancy and that end-users were viewed as a single market. As a result of these early painful experiences, the vendors began to segment the end-user market. Products, services, and market strategies were tailored to individual segments. Ease of use and variety of information were the key ingredients in the strategy for success and the foundation of a new market philosphy. Therefore, database vendors began investing in the development and application of new technology.

Which brings us to another important lesson:

Lesson Five: End-Users, even in the abstract, have considerable market leverage that benefits the entire information retrieval field.

The information systems that are evolving directly benefit information specialists as well as end-users. The continued evolution of information systems and accrual of benefits to the community will require a commitment from all involved parties.

The Challenge

As noted earlier, the number of databases in existence and accessible is increasing dramatically. This trend will continue, and we will see the addition of more "types" of databases. Numeric and factual databases will begin to number with bibliographic databases and eventually surpass them as more database vendors move toward providing full-text online.

The increase in the number of online databases means that there is more information, and more types of information, readily available. As more information becomes available online, the probability of locating needed information also increases dramatically. Consider, for example, the ease with which details from a newspaper article can be located when full-text articles are available online. And, when information cannot be located online, we can conclude with greater certainty that the subject has not yet been covered in the literature.

The challenge which arises from having so much information available is how to provide selectivity. By what measures do we ensure that the proper sources are utilized and that relevant information is retrieved? How do we avoid leaving out important sources of information and retrieving volumes of irrelevant data? Some of the difficulties associated with this task are:

- o maintaining a constant alert for new databases/services
- o registering for, and maintaining accounts with, each of the services
- o selecting the right service(s) for the query
- o learning the diverse command structures associated with each of the services
- o becoming knowledgeable regarding the terminology used in each database
- o merging and analyzing results from multiple services

Many of these problems are being resolved through the application of new information technologies.

Technology to Meet the Challenge

Recent advances in information retrieval systems toward easier and more successful use have focused on developing and applying interface technology. These advances range from the development of a simple, interactive menu for searching a single system to the incorporation of gateway and artificial intelligence technology to aid the user in navigating through numerous, diverse information systems and culling out relevant information. The end goal is to allow users to interact with these systems via an expansive natural language interface, but the technology required for this effort is not mature enough to meet the task. In the interim, technology is in place that provides users with options for executing information searches. Some of the major alternatives are described below.

Interactive Menu. The objective of interactive menu development is to design the system so straightforward that a novice user can initiate the basic operations required to perform a search. The menus remove the requirement for the user to learn obscure retrieval protocols. Often the design incorporates an option of using a limited number of natural language commands or command abbreviations as the user becomes more familiar with the system. These interfaces may be built into the information system itself or executed from a remote central processor or microcomputer system.

Common Command Language (CCL). The CCL alternative requires that the user know, or is willing to learn, at least one set of database retrieval commands. The CCL may be based on a national or international standard for database retrieval commands, or the CCL may be a system's existing commands which are translated, transparently to the user, to other systems'languages so that other systems respond to the same commands. Systems are being developed which provide the user with the choice of using a standard or an existing command structure. The CCL interface may be built into the information system itself or executed from a remote central processor or microcomputer.

Gateways. Gateways allow users to interconnect with multiple, diverse information systems. In essence, gateways act as intelligent switches. The level of intelligence is based on system design. At the lowest level of functionality, gateways handle the computer protocol and handshaking so that systems can talk to one another. Basic capabilities beyond simple interconnection include a design that relieves the user of the need to register for multiple systems, to learn telecommunication paths and protocols, and to memorize system logon and logout procedures. More advanced gateway systems incorporate menus, common command language, data analysis routines, artificial intelligence, and related technology.

Gateways can be run in environments ranging from single-user microcomputers to multi-user, central processing mainframes. The number of resources targeted for interconnection, the number of users, and the "intelligence" required in the gateway determine the configuration requirements.

Knowledge-Based Interface. Menus, common command languages, and gateways bypass significant barriers to information retrieval, but they have not broken through the human language barrier. Through the development of knowledge-based interfaces, we are inching towards the goal of natural language search interfaces. Knowledge-based systems use artificial intelligence (AI) to make the human-machine interfaces more human-like and more natural. The system performs in a manner that simulates intelligence; it makes educated decisions. To get a system to this point, we must build a knowledge base similar to that of a human expert. This requires analyzing the knowledge that the human has and the thought processes used to

make decisions. AI programming tools can then be used to program this knowledge into the system. The result is often called an "expert system."

Building a knowledge-based system is a complex task. The system is usually limited to a narrow subject scope. To be successful, the system must be programmed with a thorough knowledge of the subject area and the natural language used by human experts and their clients. Designing an interactive natural language interface is a difficult task. Most of the knowledge-based information system interfaces designed thus far relate to medical information. A cancer research file is an example. The medical field has a fairly standard vocabulary which makes it a good candidate for an expert system. An expert system with the knowledge of a librarian or information specialist is not now possible, and a development of such a system is a long-term effort at best. But overcoming the communication barrier a function at a time through use of the appropriate technology, whether menus or gateways or artificial intelligence, is essential to the continued evolution of information systems. It becomes even more critical as the volume of data we have to deal with increases.

The table below provides a sampling of interface software on the market today (Ref. 4). This list is by no means exhaustive. It notably omits non-commercial products which have been developed by the government and university communities. However, it does provide an indication of the interest and activity in the interface field.

Table 1
Representative Interface Systems

Software System	Function	User Level
BCN's Super Scout	System access to 15 services including DIALOG and Western Union's Easy Link (electronic mail)	Novice
EasyNet	System access to more than 10 services including DIALOG, SDC, BRS and NLM	Novice
IN-SEARCH PRO-SEARCH	System access to DIALOG and BRS with cross-system emulation and general telecommunications capability	Novice Expert
INSTANTCOM	Upload/Download. General telecommunications to DIALOG, BRS, Dow Jones, and other systems	Intermediate
MCI Mail	Electronic mail and system access to Dow Jones and other systems	Novice
PC/Net-Link Net Search	System access to DIALOG, SDC, NLM, Dow Jones, Lexis, OCLC, RLIN, Westlaw. Questel and Newsnet with database manager (Net Search)	Novice
PC-Talk	Upload/Download. General telecommunications capabilities	Intermediate
SciMate	System access to ISI, DIALOG, BRS and NLM with optional database manager	Novice/Intermediate
Search Helper	System access to six Information Access Co. databases on DIALOG	Novice
Searchmaster	System access to SDC, BRS, DIALOG and NLM.	Expert

CASE STUDY

The following case study describes the efforts of the U.S. Defense Technical Information Center (DTIC) to develop and implement an integrated, functional scientific and technical information network. This network was purposefully designed to incorporate both end-users and their information management intermediaries in a complementary manner making them resolute partners in the exploration of how new technology can be applied in an actual information service environment.

In 1980, the U.S. DTIC was assigned functional responsibility to develop and implement a coordinated Scientific and Technical Information Network (STINET). The purpose of the network was to facilitate DoD access to scientific and technical information relevant to DoD mission areas. Key elements of the STINET were identified as: DoD managers and scientists; DoD Libraries and Information Analysis Centers (IACs); DoD Laboratories; DTIC; and other DoD federal, commercial and foreign databases, systems and sources of STI. DTIC was tasked to provide a focus for molding the STINET, establish interoperability among the various network components, and coordinate network evolution. DTIC was to achieve this vision by promoting resource sharing and cooperative efforts and through investigation, experimentation, and application of advanced information science and technology.

Development of a STINET was a huge and challenging task which had to be accomplished within existing financial and personnel constraints. In order to make visible progress towards the STINET, DTIC had to carefully identify network requirements and evaluate the currently available technology in which applied research could be invested in both the short term and the long term to meet these requirements. Next, DTIC had to settle on an agenda of purposeful steps that would close in on the ultimate goal. Successful development and implementation of the network would depend on devising a "doable" plan of work with room for deviation when opportunities for technological acceleration became apparent.

The following elements would be required for a successful network:

- o interoperability and interconnection
- o tools such as pointers and menus, to help locate information
- o standardized command functions
- o compatible, multifunctional, flexible software for installation at network nodes
- o interconnection with diverse sources of information including government, commercial and foreign
- o interconnection with diverse types of information including numeric and factual.
- o selectivity and data analysis routines
- o improved delivery systems
- o integration of databases and people bases

The technology needed to meet the requirements of the STINET was at various stages of readiness. DTIC selected several key and promising areas in which to invest development efforts. The goal of these development efforts was to provide the basic foundation for, and functional capabilities of, the STINET. These efforts are summarized in Table 2 below.

Table 2
Scientific and Technical Information Network (STINET)
Initial Development Efforts

Development	Network Function/Description	Characteristics					
	Legend: Current ● Planned 0	Classified	Unclassified	Multi-User System	Single-User System	Stand-Alone System	Integrated Module
Gateway *Prototype: DGIS*	Provides the backbone of the network. Allows interoperability and interconnection. Interface incorporates menu and English-language commands. Accomplishes automated connections to diverse, geographically distributed resources, simultaneous searching, data uploading/downloading, and file transfer. UNIX operating system.	0	●	●	0	●	●
Local Automation Model (LAM) Prototype: IBIS	Provides the capability to create and maintain a local database/ catalog in a network-compatible manner. Unifies methodology for using internal and external resources. Provides information control mechanisms such as a circulation system. Prototype implemented using an integrated library system merged with an intelligent gateway.	●	●	●	●	●	0
Directory of Resources *Prototype: Dir. of Databases*	Provides pointers and menus to help locate information. A prototype Directory of Databases is in operation, implemented using the INGRES Database Management System. A Directory of Experts and a Directory of Computing Resources are planned.		●	●	●	0	●
Common Command Language	Provides a standardized set of commands for searching diverse databases. Implemented in C-Language for a limited set of databases. Switching to PROLOG for a production version.	0	●	●	0		●
Post-Processing	Provides a tool box of utilities for selecting and analyzing relevant data. Dependent on translating data to a common format for manipulation. Implemented in C-Language and UNIX scripts.	0	●	●	0	0	●
End-User Interface *Prototype: Search-MAESTRO*	Provides a mechanism for users who do not possess retrieval expertise to search diverse databases. Implemented using the EasyNet service from Telebase Systems, Inc.	0	●	●	0	●	●
Expert Connector Intermediary/ Expert Link	Provides a mechanism for establishing people bases or an expert network. Information experts can be linked with end-users. Implemented using the electronic mail system developed by Lawrence Livermore National Laboratory, the UNIX Link command, and the SOS capability of Telebase Systems, Inc.		●	●			●
Electronic Document System	Provides mechanisms for digital input, storage, and delivery of full text. Will improve selective delivery of information to users. Concept development stage.	C	0	0	0		

The DoD Gateway Information System, the Integrated Bibliographic Information System and the SearchMAESTRO system (described below) provide examples of how this technology is being developed, applied and integrated to provide easier and more successful information retrieval.

The DoD Gateway Information System (DGIS)

In 1983, DTIC initiated development of the DoD Gateway Information System (DGIS), an intelligent gateway system which would provide distributed networking, electronic communication, and information access and analysis. The DGIS was to link people, information services, and computers pertinent to the STINET. The technology embodied in the DGIS would provide the key menus, pointers, interoperability, and interconnection within the STINET. To accomplish this, DGIS would have to function as an electronic switch, a translator, a communications interface and a transaction controller. DGIS would require a

variety of alternatives, tailored to different user types and needs, for obtaining and distributing information.

DGIS focuses on streamlining the information retrieval and analysis process. This is accomplished by placing the user at the center of a vast information universe consisting of people bases and databases and providing the user with the navigational tools required to pinpoint and arrive at his destination. In terms of databases, the DGIS is designed to provide users with answers to the questions:

WHAT RELEVANT DATABASES EXIST?
HOW DO I ACCESS THEM?
HOW DO I RETRIEVE INFORMATION FROM THEM?
HOW DO I MANIPULATE THE RETRIEVED INFORMATION?

The DGIS provides a single, easy-to-use interface for identifying, accessing, interrogating, and post-processing information from numerous databases relevant to DoD information needs.

In terms of people bases, the DGIS is designed to answer the questions:

WHAT EXPERTISE IS AVAILABLE ON THE NETWORK?
HOW DO I COMMUNICATE WITH EXPERTS?
HOW DO I SHARE INFORMATION WITH COLLEAGUES?

The DGIS acts as an integrated information system which allows human experts, information users, and information resources to exist and interact in harmony.

Development of DGIS is a multi-year, multi-task project. A prototype system has been developed. The basic components of the system are:

A DIRECTORY OF RESOURCES, SUBJECT SEARCHABLE
A COMMON METHOD FOR ACCESSING AND SEARCHING DIVERSE DATABASES
TOOLS FOR DOWNLOADING AND POST-PROCESSING DATA
TOOLS FOR COMMUNICATING WITH A NETWORK OF EXPERTS AND COLLEAGUES.

DGIS was designed for a DoD user community including both intermediaries and end-users. Databases accessed are federal, commercial and international. In addition to large, well-known databases and systems, many small, specialized DoD databases will eventually be part of the DGIS.

Once the broad requirements for DGIS were identified, a software survey was conducted to determine if a software product already existed which would meet DGIS needs. The survey showed that THE system was NOT out there, waiting. One software package, the Technology Information System (TIS), did provide a suitable foundation on which to begin constructing DGIS (Ref. 5). TIS was under development at Lawrence Livermore National Laboratory (LLNL) under the sponsorship of the Department of Energy (DOE) (Ref. 6). TIS functioned as an intelligent gateway capable of interconnecting heterogeneous information resources at geographically distributed locations in an automated, unified, and controlled manner. In addition, TIS downloading and post-processing capabilities were already available for selected databases. Therefore, TIS was used as the baseline for developing the DGIS.

Directory of Resources

The Directory of Resources will include subdirectories with references to databases, people, and computing resources. In the first phase, the Directory of Databases is being developed; this is described below. The Directory of People will contain references to experts in subject areas and information retrieval techniques who may be contacted via the network. The Directory of Computing Resources will contain references to computing resources, such as supercomputers which can be used for data analysis and modeling, available through the network.

In order to develop a Directory of Databases, it was necessary to identify and catalog existing databases and make that information subject-searchable, so that user information needs can be matched to relevant resources (Ref. 7). Although there are many directories that identify commercial and prominent federal databases, such information was not readily available for DoD databases. To fill this void, a questionnaire was addressed to the DoD Research and Development (R&D) community to identify existing databases, their scope, and availability. Over 400 databases were identified (Ref. 8).

The next step was to build a database of the DoD and DoD-relevant databases which had been identified. A user survey was conducted to determine database requirements (Ref. 9). A database schema was developed for the Directory, and database entries were subject indexed. The database was built using the INGRES relational database management system.

The result of this effort is an online Directory of Databases which contains information on the content, scope, and availability of selected databases. The Directory is subject-searchable; upon entering a topic of interest, the user is provided with a list of appropriate databases.

Interfaces for Searching Diverse Databases

One of the primary goals of DGIS is to relieve the user of the need to learn and master separate commands and protocols for each database accessed. As mentioned earlier, the DGIS target user community includes both end-users and intermediaries. It is a rare intermediary who maintains proficiency in the use of more than five systems; for end-users, two systems is high (Ref. 10). With the ongoing proliferation of databases, it is obvious that both end-users and intermediaries will benefit from an interface for searching diverse databases. DTIC found that end-user and intermediary interface needs are very different when considered in conjunction with today's technology. An expansive natural language interface requiring artificial intelligence applications appealed to both populations, but could not be accomplished with existing technology in the short term.

A dual approach was adopted for the interface design, incorporating separate strategies for intermediaries and end-users. Eight database systems were selected for inclusion in the prototype. These included three large federal systems and two small DoD systems.

A software survey was conducted to determine if interfaces existed which met DGIS needs. One need that influenced the result of this survey was the requirement to support low-end, "dumb" terminals, as well as intelligent devices, on DGIS. Many of the software packages identified were designed for a micro-computer environment and had to be ruled out for use in the prototype. As a result of the survey, a decision was made to develop an interface for the intermediary and to integrate an existing interface for the end-user.

For the intermediary, a command translator is being developed which allows the user to interact with any of the test systems using the command language he selects. The user, for example, could search NASA/RECON using DROLS commands or the reverse. Since some commands will not have an equivalent in another database, native command searching will be retained.

To satisfy the end-user, the EasyNet database searching service has been integrated into DGIS under the name SearchMAESTRO. SearchMAESTRO is a menu-driven database front-end which provides access to over 900 commercial databases. This service was tested by members of the DoD end-user community who were delighted with the simplicity of search execution. In addition, a SearchMAESTRO interface to the DTIC Technical Report database has been developed. SearchMAESTRO access is now an option within DGIS and is described in greater detail in a later section of this paper.

Post-Processing

Information retrieved from databases often requires analysis or post-processing in order to become useful to the researcher. DOE recognized this need and developed many options for post-processing data from DOE/RECON through TIS. A library of post-processing routines for numeric and bibliographic data was available on TIS software and was incorporated in the DGIS project (Ref. 11). In order to post-process data, the user downloads it into a file on DGIS, translates the data into a common format, and calls up one of several available post-processing routines.

DTIC tested existing post-processing capabilities, developed recommendations for enhancements, and established prioritites for the expansion of the capabilities to other databases.

DGIS User Interface Design

The interface incorporated in TIS software was structured to support users with a knowledge of the UNIX operating system. As a rule, the DGIS user community lacked this knowledge and was not inclined to invest the time required to acquire UNIX expertise. However, TIS software had a flexible design which allowed the user interface to be tailored to the DGIS user community.

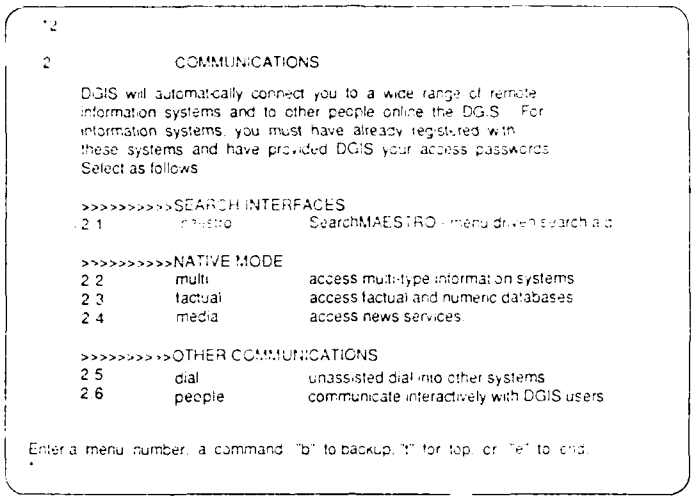
The DGIS interface incorporates menu and command modes (Ref. 12). The objective was to allow the novice user to interact with the system with ease, by descending through a series of menus. For the more experienced user, commands to execute systems functions were incorporated. The main menu of the DGIS which shows the overall processes of DGIS, is as follows:

```
0          WELCOME TO THE DoD GATEWAY INFORMATION SYSTEM

      >>>>>>>>>>INFORMATION TRANSFER MODULES
      1        directory        DGIS Directory of Online Resources.
      2        communicate      Connect to Information resources and people.
      3        process          Information product tailoring.

      >>>>>>>>>>INFORMATION UTILITIES
      4        em               Electronic Mail.
      5        files            File operations.

      >>>>>>>>>>SUPPORT INFORMATION
      6        help             Description of features.
      7        users            DGIS registered users.
      8        info             DGIS news and information.
      9        dticlog          DGIS full text retrieval.

      DGIS HOTLINE NUMBER: (703) 276-8182
      or send questions via DGIS EM to 'dgishelp.'

Enter a menu number, a command, "b" to backup, "t" for top, or "e" to end.
```

People Bases and Databases

The goal in obtaining information is to acquire knowledge. Much of the information we need resides in the minds of human experts. Therefore, the DGIS has been designed to allow interaction among people, hence the concept of people bases as well as databases.

Through the DGIS, users will be able to identify and communicate with experts and colleagues and to connect to information resources. As a first step, we have focused on providing the technology to allow such interaction. The DGIS menu for communications, shown below, illustrates how we are integrating databases and people bases. Users can access a search interface, a database/system, or people.



Accessing a database/system is accomplished using the CONNECT command or the DIAL command. The CONNECT command provides users with automatic access to information resources. Users do not have to know telephone numbers, Defense Data Network (DDN) locations, passwords, access protocols, or log-out protocols. The user enters the CONNECT command and a data resource name. DGIS then establishes a connection to the resource when the connection is made, DGIS logs the user in. DGIS uses TYMNET, TELENET, DDN, and commercial telephone lines to establish connections.

The CONNECT command can be used to access information centers worldwide. To be eligible to use the CONNECT command to access a resource, a DGIS user establishes an account with that resource and obtains the required access identification information, such as passwords, which is then programmed into the gateway by the DGIS Database Administrator. The billing process between user and resource is unaffected by gateway access. Vendors maintain the same billing structure and users maintain the same reimbursement structure, regardless of the access procedures. DGIS has several levels of security to ensure that password integrity is not violated.

As the number of resources DGIS connects to increases, we plan to have DGIS accounts that will alleviate the need for users to establish their own accounts with each vendor. To this end we are now developing a charge-back mechanism so that we can bill users for their use of these general accounts.

Users who wish to access a resource that does not have an automatic connection routine can take advantage of the DIAL command, rather than the CONNECT command. DIAL allows users to call any information center, computer, or terminal, no matter where the location. To use DIAL, the user must know the necessary passwords and telephone numbers. DIAL allows the user to access an off-network facility while retaining DGIS capabilities such as downloading and file transfer.

Once connected to a resource through DGIS, a user can download data from that resource to DGIS. Downloading data opens many options. For example, data can be reviewed at leisure, merged with other data, and shared with other users by allowing them to access the file. Data files can also be transferred to other users so that they can manipulate the data to suit their own needs. DGIS allows data to be shared selectively on a worldwide basis. The user is responsible for ensuring that copyright laws are not violated.

Several mechanisms are available for interconnecting people. An electronic mail service is available twenty-four hours a day. Standard electronic mail features such as send, receive, answer, and forward are incorporated. Mail messages can be sent simultaneously to multiple addresses and to every member of pre-established mail groups, with lengthy documents attached if needed. Users recognize the benefits of being able to communicate with numbers of people at the same time and of avoiding the "telephone-tag" routine. Messages can be filed for future reference or deleted from the system upon command.

In addition, electronic mail can be used to send information downloaded from a database and placed in a file. A user who does not want to do his own database searching can send a search request to an information specialist without leaving his desk. The information specialist can perform the search, download the results to a file, and send the file to the user. Since the data is stored in a file, both the information specialist and the end-user can use post-processing routines for manipulating and analyzing the data. The user can share the file with colleagues by using the electronic mail feature.

WRITE is another communications option which allows users online to communicate with each other via their terminals. To establish a WRITE session, the user first enters the command %WHO to get a display list of the users currently online. The user then enters the command %WRITE, followed by the name of the user with whom he wishes to communicate. The WRITE command notifies the party being called, who then has the option of responding. The command TALK is a variation of the WRITE command which provides a split screen so that both parties can input at the same time. The WRITE and TALK commands are only useful, of course, when parties who want to communicate are logged on, by chance or arrangement, at the same time.

The LINK command allows two or more users at different locations to link their terminals so that they view the same data display. All users have control over the display and can issue commands at will. Of course, linking necessitates a cooperative spirit and some coordination.

Through the LINK command information specialists and end-users can together perform interactive database searches. The end-user benefits from the specialist's expertise, while the specialist benefits from the end-user's immediate feedback.

The LINK command can prove advantageous in many other situations. For example, an instructor can provide online tutorials to a student or a class at a different location. This technique was used by Dr. Sullivan of the Chemical Information System (CIS), in Washington, DC, to provide a demonstration of CIS to a class in Brazil. The use of a speaker phone enhanced this demonstration by providing simultaneous voice communication.

The LINK command is also useful for joint online editing or reviewing of reports among experts. This practice eliminates mail delays and allows experts to discuss changes while viewing the data together.

User Support

As implementation and testing of the various DGIS modules began, it became obvious that some form of user support and training was required to ensure the success of the system. A Gateway User Support and Training Office (GUSTO) was established to satisfy this need (Ref. 13). GUSTO provides a hotline service which users can call when they have a problem. GUSTO staff will identify the source of the problem (i.e., the gateway, the user's terminal, a telecommunications link, a remote system, etc.) and take action to have the problem resolved. Users may also contact GUSTO staff using the electronic mail capability available on DGIS.

Developing user's manuals and providing training courses are also GUSTO responsibilities. The training course is primarily designed for the professional searcher who wants to exercise the full power of the system, especially in the area of post-processing of bibliographic data. The user's manual serves as a reference tool for the user. GUSTO staff also poll the DGIS user community as needed to identify new system requirements.

Prototype and Beyond

A prototype DGIS has been developed and is currently undergoing test and evaluation within the DoD community. There are currently 150 users testing the system. The DGIS prototype is running on a VAX 11/780 using the UNIX operating system, the INGRES database management system, and the PROLOG interpreter package. The DGIS software is being ported to a Pyramid 98X, an Elxsi 6800, a Gould 6050, and Sun Workstations for benchmarking and performance evaluation. Based on the results of the performance evaluation, a hardware configuration for a production system will be acquired. The production configuration may consist of several machines networked together. For example, the common command language and post-processing routines could be isolated on a backend machine. DTIC plans to stabilize a version of the DGIS and offer it as a standard DTIC service in October 1988. Prototype development will continue on a separate development machine, and enhanced versions of the DGIS will be made operational at selected intervals.

The DGIS was developed in prototype as an unclassified, minicomputer-based centralized gateway system. As we move this version into operation, we plan to begin development of a distributed, clustered gateway network. Gateway nodes in the network would be made up of centralized, multi-user configurations and intelligent workstations distributed to users. A selection of gateway capabilities would be available on personal workstations. It makes sense, for example, to have automated connection routines, common command languages, and post-processing routines for frequently-used databases available on a personal workstation. The centralized, mini-based node would be utilized to identify and search infrequently-used resources and for post-processing volumes of data which are beyond the capacity of the personal workstation. Clustered gateway nodes consisting of a centralized gateway processor and personel workstations could be based on geographic, organizational, or subject boundaries. The clustered gateway nodes would interconnect and route users to appropriate nodes when necessary. Development of a classified gateway system is also underway.

Integrated Bibliographic Information System (IBIS)

Development and implementation of the DGIS allows DTIC users to connect to, and search and analyze data retrieved from diverse unclassified database services in the federal, commercial, and international sectors. Development, as planned, of a classified version of the DGIS will make its reach almost limitless. But what the DGIS has not provided is tools for the development of local databases or catalogs of holdings and tools for local collection management.

The DGIS was designed as an intelligent switching mechanism. The resources targeted by the DGIS were already online. DGIS was designed only to switch users to automated resources. The Directory of Resources is the only database central to and created and maintained on the DGIS. This is a basic design philosophy, not to be altered for fear of deflecting the DGIS from its primary focus — that of being a gateway.

But there was a need to provide a vehicle to automate and manage local information collections which are manually maintained, very valuable, and very difficult for non-local personnel to use. This need is acutely felt by the DoD library community, a key component of STINET.

Therefore, in 1983, DTIC initiated development of a library automation system responsive to the networking and local collection management needs of DoD libraries. The system would support centralized resource sharing while allowing local processing flexibility. The objective was to permit DoD libraries to make maximum use of existing information, organize this information to meet local needs, and selectively share newly-generated information with other members of the community. The system designed to accomplish this would have to integrate local control for local collection management functions (reference, cataloging, and circulation) with access to the external resources required for reference, shared cataloging, and other network requirements.

The automation needs of the DoD libraries were defined through a requirements study initiated in 1983, which included surveys, site visits, and staff interviews throughout the DoD technical library community (Ref. 14). Based on this study, a prototype system to meet these requirements was specified, and its development became the objective of the IBIS project. The requirements that IBIS is expected to satisfy are summarized as follows.

Local Collection Management: Local cataloging, retrieval, and circulation capabilities are essential IBIS system requirements. Acquisition and serials management functions are desirable system features; they will be added at a later date.

External Database Access: System capabilities to input data to and retrieve data from external databases are critical. Uploading and downloading capabilities are essential. The DTIC TR database is a primary external resource; DTIC TR access by IBIS is vital to the cataloging and reference functions of the libraries.

Integration of Local Collection Management and External Database Access Capabilities: These capabilities are to be integrated on one computer and accessible by an authorized user from a terminal equipped with video display screen.

Common Command Set for Performing Functions Locally or Externally: A single command language is necessary for users. The IBIS will perform the necessary protocol translations between the single command language internal to IBIS and the diverse command structures of the external databases. The common command set will relieve the user of the need to learn and master separate languages and procedures for each database accessed. However, "native" language access to external databases must be available to the user.

Simultaneous Access to External Data Sources and the Local Catalog for Reference Searching: The libraries are to be able to run the same search query against multiple databases, local and external simultaneously. Search results are to be delivered to a single terminal.

Post-Processing of Retrieved Data: The ability to reformat, merge, and sort data downloaded from external sources is a desirable feature. This capability will allow libraries to fit search results -- derived from external sources and the local catalog — to their patron's needs and deliver a single product in an economical and efficient manner.

Flexible Local Catalog Format: The IBIS format must be flexible to accommodate the diverse local catalog formats used throughout the DoD community. A flexible local catalog format will encourage wide IBIS implementation. This approach avoids the delays and extraordinary expense that would be required to resolve the issue of cataloging standardization within the DoD.

Machine-Aided Citation Translation and Uploading to DTIC: The IBIS must assist in translating bibliographic citations from the local file format into the DTIC format. This capability will allow DoD libraries to contribute data to the DTIC TR database directly and efficiently. This information will then be available to the entire DoD community for display and downloading. The resulting shared cataloging will contribute significantly to meeting the resource sharing goals set by DoD.

Patron Access: IBIS will verify a patron's right to access before releasing information classified as Defense-sensitive or otherwise restricted. In addition, the IBIS will provide for online communication, as allowable by local security regulations, between patrons and the library staff and among patrons/colleagues themselves. In some instances, this will require that IBIS be connected with a local area network.

From Concept to Prototype. The design concept for the IBIS was formulated from the foregoing requirements. A software survey was conducted to determine if commercially available software packages could provide the capabilities required to implement this concept directly (Ref. 15). A list of 30 critical software functions was compiled and used to conduct the survey. Reference sources, such as the Data-Pro Reports and the Library Systems Evaluation Guide, were used to identify software packages specifically intended for application in the library environment. A total of 66 potentially suitable packages were identified, and their vendors were asked to respond to the survey. Analysis of the responses showed that three of the survey questions were key. They were:

- o Does your package support online:
  - Cataloging (to include online catalog updating)?
  - Reference/Catalog search and retrieval?
  - Circulation management and control?
  - Serials management?
  - Acquisition management?
  - Others?

- o Does your package have a gateway capability that includes:
  - Queries of "second" external database (heterogeneous)?
  - Downloading/retrieval?
  - Uploading/cataloging?
  - Post-processing?
  - Telecommunications link, auto-dialing, log-on?
  - Single query language?

- o Does your package preclude running other applications not part of your package?

The answers to the first two questions eliminated the notion that a "perfect" software system was available. No single commercial or public-domain system provided the capabilities required to implement the IBIS concept. Integrated library systems that supported cataloging and reference functions existed,

as did gateway systems that supported the querying of external databases, uploading, downloading, and simultaneous searching, but no software combined these functions. In addition, no software package could be identified that contained a suitable common retrieval language, or one which the vendor was willing to adapt to satisfy the database access requirements of the DoD library community.

It was the responses to the third question which gave hope that a tractable solution was possible. The approach chosen was to combine an integrated library system for local library collection management with a gateway system for accessing external resources. The intelligent gateway processor embedded in the DGIS was designated as the gateway portion of the IBIS. It would support access to external resources, data downloading and uploading, simultaneous searching, and post-processing. The common command language which DTIC was developing for DGIS would be made available for IBIS also.

Of the 66 library system packages identified, six were determined to have the requisite features. These six packages were benchmarked, and a selection was made in August 1985. At that time, the integration of the selected packages with the gateway software began.

The Defense Nuclear Agency (DNA) was selected as the prototype site for the IBIS. The DNA library, like many of the libraries surveyed during the requirements definition phase, had a collection of more than 100,000 holdings and required a system supporting multiple users. The IBIS prototype was implemented at DNA on a VAX 11/750 minicomputer.

Scaling Down. In the initial stages of the IBIS project, it was believed that the production system design selected as a result of the DNA prototype test would be suitable for implementations throughout DoD. However, it was determined during the software survey that most library packages are designed to operate efficiently over relatively narrow ranges of collection sizes and transaction volumes. The hardware configurations selected by the various vendors to run their different packages reflected this tendency. It then became evident that a distinct IBIS configuration to support smaller DoD libraries would be needed.

The hardware configuration required to support the prototype IBIS at DNA, a minicomputer-based system, was not economical for smaller DoD technical libraries with collection sizes ranging from 5,000 to 75,000 items. A microcomputer-based system was more appropriate for the lower transaction volumes and smaller operating budgets associated with these libraries. Therefore, DTIC initiated an effort to identify and isolate the special requirements of these smaller DoD libraries and accommodate them. A software survey was performed to identify packages which were suitable for servicing these smaller libraries and at the same time were compatible with the gateway software. The U.S. Army Training and Doctrine Command (TRADOC) library was selected as the prototype site for the microcomputer-based version of the IBIS.

Future Plans. The results and experience gained during the test and evaluation of the DNA and TRADOC prototypes will be used to develop the specifications for a competitive procurement of a production system from a commercial source. The production system will be available for purchase and installation by any library on the STINET. As a result of the dual approach -- small and large libraries -- an IBIS product line which can meet the needs of any DoD library, regardless of collection size and transaction volume, is expected to result. This product line is expected to be available to the DoD user community in the summer of 1988.

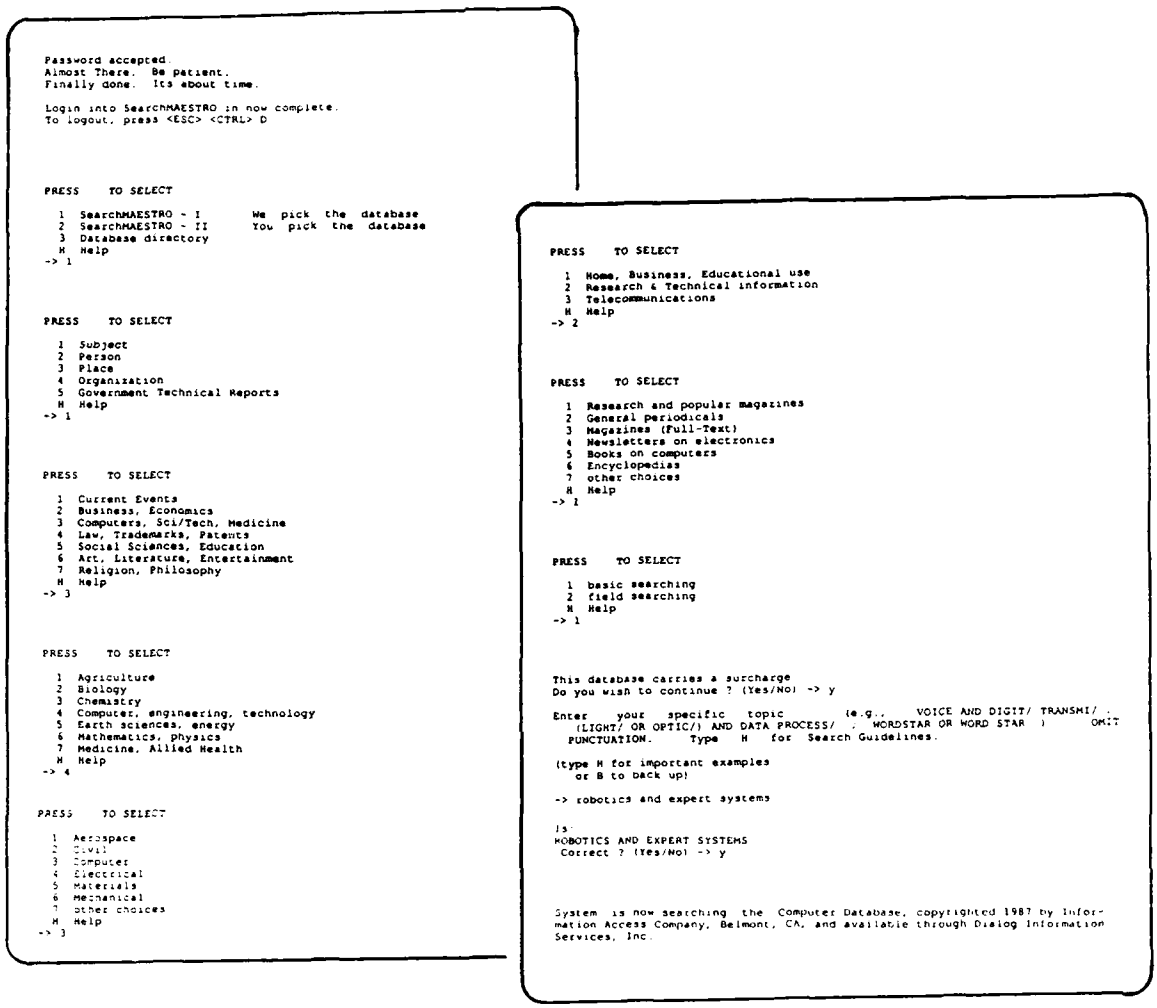
The IBIS is the first of DTIC's Local Automation Models and will make network-compatible software available for local installation and use. The IBIS is tailored for bibliographic information, and its target community is DoD libraries. Later, Local Automation Model product lines can be tailored for other user communities and different types of data. For example, models tailored for numeric data may be a requirement for some communities. Depending on the software selected for the production IBIS and the non-bibliographic user requirements identified, it may be possible to modify the IBIS software to meet additional needs. If not, DTIC can develop additional network-compatible local automation models specifically for an identified defense requirement.

Successful development and deployment of the DGIS and IBIS will provide DTIC with a powerful product/service line for the STINET. Through STINET, users would have a mechanism for interconnecting with a virtually endless range of information systems, computing resources, and people. They would also have a network-compatible vehicle for automating and managing local information, selectively sharing that information with other members of the network, and analyzing information from local and remote resources.

SearchMAESTRO

Both DGIS and IBIS were designed to deliver the power and utility required by a broad section of DTIC's market. But part of DTIC's market did not require the full power of the DGIS and the IBIS. This segment consisted of end-users who wanted to do some of their own information gathering. The users in this market segment had the requirement to scan literature and locate relevant items in their area of interest. They needed an interface that would provide easy access to a variety of databases. In some cases, the users would not be familiar with existing databases, so they required a system that would select a database for them and guide them through the search process. The information needs of these users tended towards fact retrieval -- such as the latest production statistics for a manufacturing company -- or scanning for relevant items -- such as what newspaper articles have been written on a particular subject. Their need was for some information relevant to the subject rather than an exhaustive search of the subject. We refer to them as casual end-users.

For this segment of the community, DTIC introduced SearchMAESTRO. As mentioned earlier, SearchMAESTRO provides access to over 900 databases. SearchMAESTRO can be used directly or through the DGIS. Users who access SearchMAESTRO directly can access databases via a simple-to-use interface. This interface eliminates the need to learn unique database command languages and search techniques. The user has two modes of operation from which to choose. In the first mode, SearchMAESTRO leads the user through a series of questions and answers and selects the database for him. An example of how this works is:



In the second mode, the user can select the database he wants and use the SearchMAESTRO interface for executing the search. With either mode, the user can view search results on the screen, or print or save the results using local equipment. The main reason for accessing SearchMAESTRO through the DGIS is to take advantage of the post-processing and electronic mail utilities. The users must always observe fair use practices when dealing with copyrighted material.

The unclassified portions of the Technical Reports and the Work Unit Information System files of DTIC's Defense Research Development, Test and Evaluation Online System (DROLS) are being made available to registered users through SearchMAESTRO.

Any time man and computer meet, a diversity of problems arise which can best be handled through human intervention. Therefore, SearchMAESTRO provides online user assistance through a function called "SOS" for "Save Our Search." At any point during a SearchMAESTRO session, the user can simply enter "SOS" and a search expert will respond. The search experts are trained to interpret reference questions, be knowledgeable about available sources of online information, and know how SearchMAESTRO works. DTIC supplies the search experts for the DROLS files, and Telebase Systems, Inc., supplies experts for all other systems.

SearchMAESTRO is currently available as a prototype to approximately 30 users. These users access SearchMAESTRO both directly and through the DGIS. The SearchMAESTRO service will be offered as a DTIC production service in October 1987.

SearchMAESTRO is just the first of several end-user interface options we plan to offer our users. DTIC has sponsored two conferences in the area of interface technology and maintains a constant alert for new offerings (Ref. 16, Ref. 17). We are especially interested in expert system and personal computer based interfaces.

Case Conclusions

Implementation of the prototypes described in this case study demonstrates the feasibility of integrating diverse, yet functionally compatible, automation and information resources within an interoperable network. It also demonstrates that a variety of technologies exist today -- software, hardware, and telecommunications -- which can be blended and modified to make using information retrieval systems easier and the results more meaningful. But technology alone is not the answer. Advances in information

retrieval systems are dependent on people — end-users, intermediaries, vendors, producers — working together and willing to modify and blend their skills in new ways.

DTIC is planning for the evolution of the STINET from prototype applications to a fully functional, integrated network. It has been said that "the best plan is only a plan ... unless it degenerates into work" (Ref. 18). DTIC is positioning itself to take advantage of evolving technologies to forward STINET development, but there are no quick technological solutions in STINET. DTIC must work to build-on, tailor, and accelerate the introduction of new technology. At the same time, we must work to effect changes in human processes so that the fullest benefit can be reaped from application of new information technologies. The STINET plan has degenerated into work! The STINET work is of a truly synergistic nature in that the whole is much greater than the individual technologies, and STINET will only come into being through the combined efforts of many people.

FORECAST

Information leads to knowledge, and knowledge has always meant power, wealth, and control. Technology is reshaping the ways we communicate, distribute, and manage information. It works to bring information closer to its users. Introduction of technologies such as personal computers and user friendly search tools is creating a massive new market for information retrieval systems. During the next five years, there will be a restructuring of the products and a repositioning of the players in the information retrieval market.

Easy-to-use interfaces are exposing new users to the wealth of information available through information retrieval systems. As they earn learn to use its power, the perceived value of information will grow. As the amount of information continues to explode, so will the demand for information tailored to user needs. Gateway systems will evolve which present users with "views" of information personalized to meet their needs. Information "boutiques," organized by subject and interconnected, will form a virtual world-wide, multi-media library.

Users want answers! This will be the overriding challenge facing the information retrieval market. Decision support systems that incorporate current knowledge and can aid users in adding to that knowledge base will be the goal. As the generation of information continues unimpeded, users will demand sophisticated information analysis tools and instantaneous fact retrieval. This will dictate that high-value information be available in digital form so that users can rapidly access facts and analyze data.

Local systems for creating, storing and processing information will become a vital part of the information retrieval market. Selected information will be distributed on optical media for local use. Information on optical media will often be part of a complementary product line which includes online services. Users will be connected to local area networks that will encourage information sharing. This will swell the information flow and make it even more difficult to pick out facts or trends. It will also increase the need for information specialists and experts.

Information specialists will be the gurus of the new age. They will provide expertise in the politics of information and will be invaluable corporate resources. They will have to respond to users' critical needs for fast, accurate, complete information. They must be throughly knowledgeable about the sources of information, both corporate and external. They must understand emerging information technology well enough to offer a range of solutions to meet the needs of their clientele. This may include building "views" of information to meet individual needs, tailoring interfaces to allow clientele to retrieve information from their "views" easily, and performing comprehensive searches for clients when exhaustive recall is essential. Most importantly, information specialists will be called upon to attest to the validity and reliability of data sources. They will have to ensure that external resources they use, or recommend to their clients, are credible. They will have to develop policies concerning how locally generated information will be controlled and validated. This will be critical; the use of incorrect information for decision making and planning could prove disasterous to an individual or to an institution.

The evolution of optical disk storage, computers and interface software provides the opportunity for rapid advances in information retrieval systems, but we are still a long way from having the world of information on a microchip at our fingertips. The information market will evolve in that direction as end-users, information specialists, information producers and vendors continue their partnership in the quest for knowledge.

REFERENCES

1. Williams, M.E., Rouse, S.N., Computer-Readable Bibliographic Data Bases - A Directory and Data Sourcebook. 1976, ASIS Publications, Chicago, IL.

2. Williams, M.E., Computer Readable Databases - A Directory and Data Sourcebook. 1985, American Library Association, Chicago, IL.

3. Williams, M.E., "Highlights of the Online Database Field - Gateways, Front Ends and Intermediary Systems," National Online Meeting Proceedings. May 1985, Learned Information, Inc., Medford, NJ.

4. Burton, H.D., The Intelligent Gateway: A Dynamic Resource Environment. July 1986, Lawrence Livermore National Laboratory, Livermore, CA.

5. Cotter, G.A., An Intelligent Gateway for the Department of Defense: The Technology Information System. June 1984, ADA-133-800, National Technical Information Service, Springfield, VA.

6. Hampel, V.E., Bailey, C., Kawin, R.A., "TIS" An Intelligent Gateway Computer for Information and Modeling Networks. August 1983, ADA-135-916, National Technical Information Service, Springfield, VA.

7. Jacobson, C.E., Cotter, G.A., The DoD Gateway Information System (DGIS) Directory of Resources. February 1986, ADA-185-244, National Technical Information Service, Springfield, VA.

8. Jacobson, C.E., Cohen, R.S., Michel, J.T., Directory of DoD R&D Data Bases. September 1984, ADB-085-600, Defense Technical Information Center, Alexandria, VA.

9. Chastain, G.C., A Study of User-Defined Searching Requirments for the On-Line Version of the Directory of DoD Sponsored R&D Data Bases on the Defense Gateway Compter System. March 1985, ADA-153-000, National Technical Information Service, Springfield, VA.

10. Williams, M.E., "Highlights of the Online Database Field - Gateways, Front Ends and Intermediary Systems," National Online Meeting Proceedings. May 1985, Learned Information, Inc., Medford, NJ.

11. Cotter, G.A., The DoD Gateway Information System. October 1985, ADA-161-701, National Technical Information Service, Springfield, VA.

12. Kuhn, A.D., Cotter, G.A., The DoD Gateway Information System (DGIS): Interface Design. March 1986, ADA-174-150, National Technical Information Service, Springfield, VA.

13. Powell, M.E., GUSTO: Training Tailored to User Needs. May 1987, ADA-181-105, National Technical Information Service, Springfield, VA.

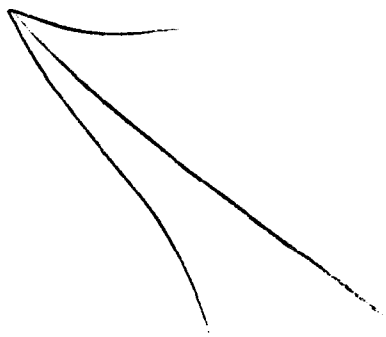
14. Hamilton, W.P., Hartt, R.W., O'Connor, D.J., Local Automation Model: Functional Definition. October 1983, ADA-133-389, National Technical Information Service, Springfield, VA.

15. Hamilton, W.P., Hartt, R.W., O'Connor, D. J., Local Automation Model: Assessment of Library Software Availablility. September 1984, ADB-087-513, Defense Technical Information Center, Alexandria, VA.

16. Powell, M.E., Selected Papers from the First Conference on Computer Interfaces and Intermediaries for Information Retrieval. October 1984, ADA-167-700, National Technical Information Service, Springfield, VA.

17. Jacobson, C.E. Proceedings of the Second Conference on Computer Interfaces and Intermediaries for Information Retrieval. May 1986, ADA-174-000, National Technical Information Service, Springfield, VA.

18. Drucker, P.F., Management: Tasks, Responsibilities, Practices. 1974, Harper & Row, New York, NY.



INFORMATION TECHNOLOGY TO FACILITATE GROUP INTERACTION,

ARCOvision

Cynthia A. Savage
Supervisor, ARCOvision Services
Atlantic Richfield Company
1500 Market Street
Philadelphia, PA. 19102

Atlantic Richfield has been successfully operating an interactive, full-motion video teleconferencing system since September of 1983. The system, named ARCOvision, is for internal use only and consists of six locations in the United States including Alaska. The implementation, capabilities, equipment, and costs related to the ARCOvision system will be discussed in this paper.

In September of 1983 Atlantic Richfield introduced a live full-motion video-conferencing system to employees in Anchorage, Alaska; Dallas and Houston, Texas; Los Angeles, California; Philadelphia, Pennsylvania and Washington, D. C. Using its own satellite-based communications network to transmit full-motion video images between company headquarters, the meeting facilities allow ARCO employees the opportunity to meet their distant colleagues face-to-face without leaving their office building. The system has been named ARCOvision allows participants to view color life-size images of all their counterparts at one time - as if they were sitting across the conference table rather than across the country.

The ARCOvision system was made possible by an enhancement of the ARCOnet system, the backbone of ARCO's communication system. ARCOnet is one of the largest privately owned and operated telecommunications systems. As part of the ARCOnet enhancement we participated in the building of four and have leased two satellite earth stations. We have leased a full transponder on Alascom's Aurora satellite. The system is completely digital and uses one of the most sophisticated multiplexing and polling techniques developed. It supports all of our voice, data, and video transmissions. The all-digital system gave us capabilities that were never economical before. The digital system gives us the ability to transmit video through very sophisticated compression algorithms and hardware.

The first question often asked is "Why do we use videoconferencing?" In our case, ARCOvision came about because our former Chairman of the Board, Mr. R. O. Anderson, told Business Week that there had to be a better way for our business to communicate. In 1980 he stated that ARCO would lead the way with sophisticated videoconferencing. In order to give you an idea of why we went into this some facts are relevant; In 1980, the year we had several tens of thousands employees, ARCO spent some 60 million dollars on travel and had more than 80 thousand business trips. While the cost in actual transportation dollars is high, the cost in human terms is even higher. Booz Allen and Hamilton did a study of business practices in America and the results indicated that there was a one-to-four productivity loss ratio related to business travel. This means that for every hour spent in a meeting to which one travels, on the average, there were four hours spent in just the logistics in getting there.

Once the decision to institute videoconferencing was made, ARCO conducted a survey of more than 1000 of its managers and professionals asking them how they perceived their needs for videoconferencing. The survey was conducted by the CCRG group from USC Annenberg School of Communications. They came with a prepared questionnaire. It sought to profile the user communities, communications patterns, meeting habits, preconceptions, business attitudes by the company, city and job classifications. As it turned out it captured our culture. Also as a part of the USC study, it was determined the key-operation cities that would be linked by this system. From that survey came the concept of what ARCOvision and the conference room must be. We knew at that time that it needed to be a full-motion, life size, open mike system. It must allow eye contact and have several presentation capabilities. The facility had to have full-screen capabilities, cameras, mikes, chalkboard capability, lighting, graphics panels, and the ability to project slides.

As far as we know, ARCOvision, was the first long-distance system tailored specifically to the needs and culture of our company. In addition, it was the first nationwide in-house network. The rooms are equipped with the full range of graphics materials typically used in ARCO business meetings - slides, viewgraphs, and chalkboards. Each identical ARCOvision room is comfortable and easy to use. Participants appear life size in full color with clear audio transmission. The computer controlled color cameras permit close examination of charts, slides, or any other object in the room while still maintaining a presence of the participants. The syst m allows eye contact and has several presentation capabilities. For example, the multiple video display options afford users either a wide screen view of all participants in the remote location or a voice-actuated one shot of just the person speaking in the other room. The advantage of the wide screen view is that participants never lose contact with each other and they are able to pick up visual cues through body language as they would in an in-person meeting. High speed facsimile hook-ups between

all ARCOvision rooms make it easy to distribute charts, text, or graphics to the other location. The system has very simple controls. It is very user friendly and allows our people, with a minimum amount of training, to feel very comfortable with the operation. We found that if one keeps the system very friendly, and if there is a real dialog during the videoconference, that after about five minutes people forget all about the technology. Because up to ten people at each location can participate in the videoconference the entire project team can attend the meeting increasing productivity and saving time.

Our ARCOvision rooms are used by all levels of management and require secure transmission for highly sensitive material. The room is isolated; both the audio and video signals are scrambled before they leave the room and again when they reach the satellite earth station. The word "codec" can be broken down into its two component words, code and decode. The coding or compression takes place according to the algorithm chosen requiring less bandwidth for cost efficiency. An added feature is the ability to enter a specially chosen alpha/numeric code, an encryption code, at the sending and receiving location only. A particularly cautious user may even set his own encryption code moments before the meeting starts to be certain his communication cannot be intercepted. The signal is encrypted and scrambled before it leaves the conference location on the way to the earth station. Once the signal reaches the earth station it is scrambled again before being transmitted to the satellite. We believe our system is very secure and it is being used for high level negotiations.

Our system is four years old; it's expanded to include 23 U.S. cities plus Canada and Europe. We arrived at this expansion not by building additional rooms for our private use, but by negotiating with AT&T and the Meeting Channel. There are three ingredients required for videoconferencing - some way of presenting the signal, a Codec to compress the signal and a means to transmit the signal. Compression of the signal requires the use of less bandwidth making the concept cost effective. There are a variety of compression equipment; codecs, available. Three of the codecs that ARCO uses are NEC (Nippon Electric Company), GEC Video Systems, and CLI (Compression Labs Incorporated). ARCO chose to use the Nippon Electric Company (NEC) codec and all of the ARCOvision rooms are equipped with NEC codecs. The problem encountered is that the different codecs manufactured are not able to communicate with one another. Therefore unless we use the same codec our video-conference centers could not communicate with yours. At ARCO we have overcome this communication problem two different ways. An additional codec, a GEC codec, was purchased and is installed in the Philadelphia room. This makes the Philadelphia location a gateway for any additional services, usually international, that use a GEC codec. For meetings originating in Philadelphia the signal is routed to the GEC codec. For meetings originating in one of the other ARCOvision locations the signal is routed to the Philadelphia room and then enters the GEC codec to the international destination. (The Philadelphia room also has access to AT&T Accunet Reserve system that serves as the transmission medium for international meetings.)

U.S. Sprint and their conference centers called "The Meeting Channel" use the CLI codec and in this case the Los Angeles room acts as a gateway. Differently in this case, the signal is transmitted by wire from the ARCOvision room on the 37th floor to the Meeting Channel room in the lower concourse of the same building. There it is put into the CLI codec and transmitted to the chosen location on the Meeting Channel network. The Meeting Channel provides connections to Japan and Canada.

The conference locations listed above cover many time zones. A system had to be implemented so that the requested time would translate to the proper local time. This was overcome with the use of Greenwich Mean Time (GMT), the prime basis of standard time throughout the world. The use of Greenwich Mean Time enables us to schedule conferences and it translates the time to the correct local time no matter what time zone.

Training is also key. When people are interested in using the system, our conference coordinator will provide them with an opportunity to experiment, learn the controls, and actually talk to another city just to get a feel of how the system works. Each ARCOvision room is run by a full-time coordinator, whose duties include scheduling the facilities, customer training, administrative support, equipment maintenance, first-level technical troubleshooting, and internal marketing. We estimate that these employees spent 60% of their time on daily operations; 20% on maintenance; 10% on marketing and 10% on administrative duties.

In addition, since this is an in-house system, we have a Network Operations Center located in Los Angeles, which is staffed 24 hours a day and which provides technical services for the satellite system, both to our videoconferencing staff as well as other customers leasing time from ARCO for their voice and data transmissions.

As for ARCOvision, we spent some time evaluating the cost/benefit trade-offs of videoconferencing and found that we had a very favorable return if only 15% of the travel were reduced between cities served by the new system. In 1980, the baseline year for our study, we spent more than 60 million dollars on 52,000 trips to various company locations. ARCOvision was planned to serve the cities which accounted for more than half that expense. So you see, relative to the "enormous costs" of travel, ARCOvision is quite economical.

It should be noted that we never intended the system to replace all business travel. In general, we expected it to supplement, rather than supplant, necessary business trips. However, we also believe that there are many meetings which could be conducted as effectively using this medium as using the old - an airplane trip - and have found that to be the case.

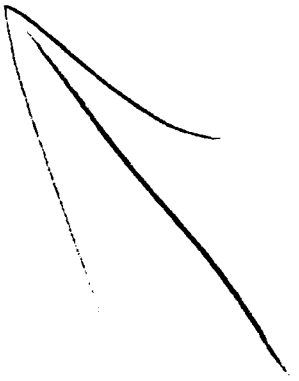
Some of the intangible benefits of ARCOvision is better utilization of manpower and quicker decision making. The ability to bring more people easily together to be involved in the decision making and information sharing process builds better morale and team spirit.

In 1986, we saved something approaching 2500 work days. This translates into a man-day savings of about $700,000. In the same year ARCOvision experienced 790 system hours of use and a total travel savings of approximately $1,300,000. The combined savings through the use of ARCOvision is approximately $2 million. This benefit, while dramatic, really does not capture all of the true benefits of a system like this. We charge, internally $480.00 an hour for the usage of the system. When we use it properly we estimate we are saving two thousand dollars an hour.

The cost of the network is justified on a cost-avoidance basis - we save about $8 million a year by having our own communications facilities. In addition, we have the advantages of security and control. The first year of existence there were no charge-backs. The second year we started charging $350 per hour and the third year $480 per hour. This does not recoup our operational costs for the system, which are at about $650 per hour; however it does increase cost awareness by our employees.

In addition to our internal costs, there are those associated with our interconnect to international locations and the Meeting Channel. Our corporation has negotiated reduced rates with these organizations. Costs are being figured on an as-needed basis; as both of these organizations are in the process of reducing their rates. However, as an example, if we did a videoconference from Houston to Paris the cost would be in the neighborhood of $2500 per hour. An Anchorage to San Francisco would be about $700 per hour.

We are proud of our videoconferencing system. We really believe it provides benefits in terms of productivity, in terms of improved participation, in terms of shortening the decision-making cycle. And it saves dollars. All of the things we set out to do, we feel we have accomplished. We believe that the original vision provided by Mr. R. O. Anderson has been taken into account. As a company, we're going the distance in providing tomorrow's decisions today. Videoconferencing helps accomplish that.



WORDS: KEYS OR BARRIERS TO INFORMATION TRANSFER

by

Mrs Suzanne Richer
Chief Librarian
Secretary of State
Ottawa, Canada

Abstract

Terms can create or resolve barriers in the processes pertaining to the transfer of knowledge and the access to information. Selecting words to organize, process and retrieve data leads to the establishment of hierarchies. The building of thesauri provides for tools enabling the organization of concepts. National and international guidelines have been developed for the establishment of monolingual and multilingual thesauri.

New words find their way in solid hierarchies. The introduction mecanisms for new words rely on cross-referencing to already established and recognized terms. Manual and machine indexing provide the same basic tools, machine indexing being, in many cases, more feasible.

Standardized terms are compiled in special listings such as the glossaries prepared and published by NATO. The linguistic quality norms of standardized terminologies are limited to the specialized objectives of such terminologies.

Terminological barriers are costly and should therefore be reduced in order not to hinder the transfer of knowledge and access to information data.

Words: Keys or Barriers to Information Transfer

History shows that man experienced information transfer barriers as early as he began organizing knowledge. The arrangement of the first encyclopaedias produced in the 17th century, in France and in England, illustrate this fact. Larousse inclined towards the alphabetical arrangement gave in, in fact, to his predecessors who aimed at classifying knowledge and the concepts and words relating to the realities of their time according to broad categories. The popularity of the alphabetical dictionary and encyclopedia favored the recognition of new words; updating a reference tool such as an alphabetical dictionary became a much easier task and proved also to be a good investment. Quantity and quality were not necessarily interlocked criteria.

The word "thesaurus" was first popularized by Peter Mark Roget with the publication, in 1852, of his Thesaurus of English Words and Phrases. Organized in terms of words related to concepts, his thesaurus pointed the way for all modern thesauri. The adaptation of a thesaurus to an information retrieval system was first referred to in an IBM journal in 1957 whereas the first thesaurus actually created for controlling the vocabulary of an information retrieval system was developed around 1959 by the DuPont Company.

In 1962, the Educational Resources Information Center (ERIC) developed an information retrieval system requiring the building of a huge thesaurus. In 1966, guidelines and conventions managing the selection of appropriate technical subject indexing and retrieval terminology as well as the display of this terminology in a thesaurus format were outlined in the US Department of Defence Manual for Building a Technical Thesaurus. These conventions were developed by Project LEX, the DOD Technical Thesaurus Task Force, in cooperation with the Engineers Joint Council. This Manual focuses on the criteria for determining the usefulness of prospective thesaurus terms or descriptors, on the means of resolving ambiguities among descriptors and on the methods for systematically creating a cross reference structure that would display hierarchical and conceptual interrelationships.

Most thesaurus formats are developed according to basic organization rules governing not only the corpus but most important the access keys allowing for the indexing or classification of concepts and essentially the retrieval of the data or information processed according to the thesaurus. Basically, all thesauri should comprise an introductory statement explaining the purpose of the thesaurus and its arrangement, an alphabetical listing of all descriptors and cross references, a corpus of descriptors arranged by subject categories based on a subject category list, a graphic display of hierarchical relationships and finally a permuted display of all descriptors in the order of each meaningfull word. These structural elements garantee the quality of the thesaurus system and minimize the linguistics and terminological barriers with regard to information processing and transfer of knowledge.

In 1970, an International Conference on the General Principles of Thesauri Building was held in Warsaw under the auspices of the Polish Academy of Sciences. The aim of

the conference was to hold a working discussion leading to the identification of problems relevant to the building of thesauri in order to develop a uniform comprehension of this building process to ensure easier communication and exchange of information between information systems covering different ranges of subject matter, and using different languages. Discussions centered on the meanings of terms such as "thesaurus", "descriptor" and "ascriptor"; the participants also focused on the fundamental elements of thesaurus arrangement and special attention was given to the methodology of thesaurus building, the selection and quality of descriptors and thus the interrelations. This international gathering of experts in thesaurus building also assessed the problems relevant to the building of monolingual thesauri and polylingual thesauri.

Unesco's Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri for Information Retrieval, published in 1970 emphasizes the needs for standards in this activity.

Unesco also published, in collaboration with ISO, the Guidelines for the Establishment and Development of Multilingual Thesauri in 1976 and revised in 1980. This standard reviews the new kinds of problem posed by multilingualism most of which are concerned with the establishment of equivalences between terms in different languages and allows for a range of options for problem solving based on the fact that a standard for multilingual thesauri can never be prospective.

ASTUTE (Automated System for Thesaurus Updating, Testing and Editing) was developed, 10 years ago, to produce multilingual thesauri for EURONET. Those already published thesauri have proven most efficient in machine translation.

The evolution of thesaurus building comprises the use of technological tools. Thesaurus construction relies on mini-computer applications in many cases. The mini-computer is used to collect, edit and correct candidate thesaurus terms and therefore eases the process of grouping terms into files of similar concepts and facilitates the generation of products useful in vocabulary review and in term structuring. Because thesaurus terms are already in machine-readable form, it is simple to prepare print programmes to provide permuted, alphabetic, hierarchical and chart formatted term displays. The use of the mini-computer facilitates initial thesaurus entry development by reducing clerical effort, editorial staff decisions and overall processing time.

Classification schemes, subject headings lists and thesauri are tools developed to control vocabularies of information storage and retrieval systems. There are many basic features that are common to both classification schemes and thesauri. The major difference is that in thesauri, words denote both concepts and ranking whereas in classification schemes, notation denotes ranking.

The selection of descriptors for both thesauri and classification schemes building has arouse many theories among librarians, information specialists and terminologists. There are basically two different strategies for identifying candidate terms. A consensus of experts of a field (the committee approach) is different from the knowledge expressed in the literature in this field (emperical approach). A number of researchers have studied the validity of this assumption only to conclude that the two strategies failed to generate significantly different lists of terms and thus raising important questions for those involved in establishing guidelines for thesaurus design and foremost for those implementing the guidelines.

In automatic indexing, the frequency characteristics of terms in the documents of a collection have been used as indicators of term importance for content analysis and indexing purposes. In particular very rare or very frequent terms are normally believed to be less effective than medium-frequency terms. More recent automatic indexing theories have been devised to use not only the term frequency characteristics but also the relevance properties of the terms. The term precision and term utility weights that are based on the occurrence characteristics of the terms relevant (as opposed to the nonrelevant) documents of a collection are introduced and methods are suggested for estimating the relevance properties of the terms based on their overall occurrence characteristics in the collection. An evaluation assessing the results of the weighting systems using the term relevance properties with the more conventional frequency-based methodology determines the measurement of term importance in automatic indexing systems.

At present, three kinds of automatic indexing systems are available: (1) inverted file full text indexing, (2) dictionary-based indexing and (3) dictionary and rule-based indexing. The only method which is fully automatic is still the full text indexing.

In 1982, the British Library conducted a study of free language vocabulary analysis especially oriented towards the automatic generation of indexing elements for document and query representation and their effectiveness in mechanized information retrieval. Automatically derived sets of index terms were evaluated for their performance in computerized retrieval queries and conclusions confirm that performance of searches on the automatically generated free index terms is comparable to searches on manually-assigned free index terms in the low recall-high precision range. High recall was possible with the automatically assigned subject headings and classification but with poor precision. Cost studies showed that the automatic system for generating indexing terms was less costly than the manual system.

In selecting terms, a review of all candidate descriptors must be conducted to avoid proliferation, maintain consistency and clarify any potential ambiguities with already established and accepted terms. It is therefore not a simple exercise to justify new terms.

At this stage, a distinction should be made between "terms" and "identifiers" or "names" such as specific projects (Brave Lion, "Big Lift" Operation), geographic locations (Camp David, SACLANT (Supreme Allied Commander Atlantic), Anti-Submarine Warfare (ASW) Research Centre), persons (Secretary General NATO Supreme Allied Commander), trade names (SIDEWINDER (nuclear missile)), programmes (NATO Sea Sparrow, NATO Patriot Programme), legislation (Brussels Treaty, Vandenberg Resolution), organizations (Allied Communication Security Agency, Central Europe Pipeline Office), equipment (air-delivered weapons, F-16 air combat fighter), acronyms (NATO, AGARD).

New terms should be added to solid hierarchies, logical synonyms and contemporary terminology. Terminological barriers are avoided, possibly not solved, by cross-referencing and linking new terms to established and recognized hierarchies. A continuous updating and sorting process adds to the feasibility and efficiency of this approach, especially as far as managing a thesaurus and a terminology database. The late introduction of new terms tends to favour the creation of parallel and ambiguous terminologies. This again can be avoided by cross referencing and linking new terms to existing hierarchies.

A new term should always be fully documented according to reliable documentary and bibliographic sources in order to avoid ambiguities. The term should also be analyzed in its proper context in order to link it to the category of hierarchy to which it belongs. At this point, cross-referencing should be used to clarify interrelations, inclusions and exclusions.

In military terminology, "a service truck" is just a truck whereas from a civilian point of view "a service truck" is a truck equipped for emergencies and routine servicing used, for example, by telephone and hydro companies and others.

Proper management of specialized terminologies through glossaries, vocabularies and other lists of terms, maintained by either manual or computerized means, offer an economical and feasible approach to the reduction of terminological barriers. One of the better known examples is without any doubt the Dictionary of Military and Associated Terms published by the Joint Chiefs of Staff, in 1979, in Washington.

DOD dictionaries and glossaries (as well as NATO's) are published for mandatory use by DOD and NATO management. Terms which are identified as DOD, for example, are not otherwise defined, unless a different application is intended. Definitions which identify a different application are to be submitted to a terminology committee for consideration and approval. In the Department of Defense Dictionary of Military and Associated Terms, terms and definitions identified with "NATO" and "CENTO", represent the US official position in international standardization agreements and in order to ensure the integrity of these commitments, terms so designated are to be used by US members of working parties, groups or panels of experts in the processing of all standardization agreements. When an agreed international term does not exist, the DOD term will be used in establishing the United States position.

The standardization mandate regulates also the production of separate dictionaries and glossaries containing terms and definitions for joint usage which cannot be published within DOD without prior coordination and approval by the authorized officers.

An extremely important note comments on the editorial objectives of the DOD dictionary by simply indicating that "... the DOD dictionary is not to be used as an editorial and style guide...".

Standardized terminologies follow the same type of processing. It is important to underline, at this point, that standardized terminologies are not necessarily of high linguistic quality. The objectives set by term banks of standardized terms allow for limited linguistic quality unless the standards from which the terms are extracted were prepared and written by teams of experts, cautious of both technological and linguistic norms.

Specialists and experts create new terms reflecting the advancement of their discipline and field of research. These new words make their ways in papers presented at scientific and technical conferences and in publications. Once created and fed into a network, regional, national or international, those new words are picked up, used and disseminated.

Whether the new words are linguistically good according to established linguistic quality criteria, they are used and therefore exist. In recording new terms, one has to establish a profile: history and genesis of the new word, definition and context, sources quoting the new words, interrelations to existing terms, comparison of definitions, contexts, synonyms, etc. The terminological record has to be documented and exhaustive.

(Examples extracted from The North Atlantic Treaty Organizations: Facts and Figures. Brussels, NATO Information Service, 1981).

Terminological barriers will always hinder the transfer of knowledge and the access to information. Terminological barriers are costly; their impact on communication systems of various nature has to be taken into account. The Tower of Babel could have easily been the first term bank.

Bibliography

European Congress on Information Systems and Networks. Third European Congress on Information Systems and Networks: Overcoming the Language Barrier, v.1: Progress Towards Standard Guidelines for the Construction of Multilingual Thesauri. (Luxembourg: 3-6 May 1977) By Derek Austin. 2nd ed. München, DE: Verlag Dokumentation; London, GB: K.G. Saur, 1977, p. 341-402.

Freeman, Frank. "Building a Thesaurus for a Diffuse Subject Area. In Special Libraries. v. 67, no. 4, Apr. 1976, p. 220-222.

Ghose, Amitabha and Anand S. Dhawle. "Problems of Thesaurus Construction." In Journal of the American Society for Information Science. v. 28, no. 4, July 1977, p. 211-217.

Harding, P. Automatic Indexing and Classification for Mechanised Information Retrieval. London, GB: British Library, Research and Development Department, 1982, 109 p. (BLRD Report, 5723)

Iljon, A. "ASTUTE: A Software Package for the Creation, Production, Publication and Printing of Monolingual and Multilingual Thesauri/Le logiciel ASTUTE: un outil informatique pour la création, la mise à jour, l'édition et l'impression de thésauri mono et multilingues." In Documentaliste. v. 14, no. 5-6, Sept.-Dec. 1977, p. 25-33.

Kazlauskas, Edward John and Thomas D. Holt. "The Application of a Minicomputer to Thesaurus Construction." In Journal of the American Society for Information Science. v. 31, no. 5, Sept. 1980, p. 363-368.

Keitz, Wolfgang von. "Automatic Indexing and the Dissemination of Information." In Inspel. v. 20, no. 1, 1986, p. 47-67.

Proceedings of the International Conference on General Principles of Thesauri Building. (Warsaw, PL: 23-27 Mar. 1970) Warsaw, PL: Polish Academy of Sciences Documentation and Scientific Information Centre, 1970, 191 p.

Salton, G. H. Wu and C.T. Yu. "The Measurement of Term Importance in Automatic Indexing." In Journal of the American Society for Information Science. v. 32, no. 3, May 1981, p. 175-186.

Seminar on Thesaurus in Information Systems. Principles for Thesaurus Construction, Sec. A. (Bangalore, IN: 1-5 Dec. 1975) Jointly organised by DRTC and INSDOC. Bangalore, IN: Documentation Research and Training Centre, 1975?, 87 p.

Trester, Delmer J. ERIC: The First 15 Years: A History of the Educational Resources Information Center. Washington: Educational Resources Information Center, 1979, 385 p.

LA BARRIERE LINGUISTIQUE : PROBLEMES DE TRADUCTION

I. CLEMENS, Ir. civil électr., Chef du Centre Belge de Traduction, CNDST-CBT, 4 Bld de l'Empereur B-1000 Bruxelles, Belgique.

RESUME

Les problèmes linguistiques sont étroitement liés aux problèmes d'informations. Ils sont donc examinés conjointement afin de résoudre les problèmes terminologiques et d'accéder à la masse documentaire. On examine les problèmes de compréhension des textes par le spécialiste et les difficultés rencontrées par le traducteur. Les problèmes linguistiques sont examinés en fonction des particularités des pays unilingues et multilingues et en fonction des groupes de langues. Les moyens existants pour réduire la barrière linguistique et les institutions impliquées sont considérés du point de vue de l'utilisateur et du traducteur. Le traducteur est le personnage-clé pour l'abolition de la barrière linguistique. Il est confronté aux problèmes de compréhension de textes scientifiques et techniques, aux problèmes de recherche de sources d'information, de termes techniques ainsi qu'aux problèmes d'adaptation aux méthodes modernes de travail. L'emploi de l'informatique lui procure une aide certaine et de multiples possibilités de rationalisation du processus de traduction. Parmis les modes de traductions connus, la traduction assistée par ordinateur paraît actuellement la plus performante : utilisation des PC pour le processus de traduction proprement dite, emploi de dictionnaires informatisés et de bases de données bibliographiques en tant que sources terminologiques multiilingues fiables et actualisées. L'ensemble des mesures évoquées permet de réduire les difficultés linguistiques du consommateur de l'information et d'assurer une aide efficace au traducteur.

I. La barrière linguistique et l'information scientifique et technique moderne

1. Obtention des documents en texte intégral et extraction de l'information
   a) disponibilité des documents
   b) contacts avec les détenteurs d'information
2. Problèmes linguistiques
   a) problèmes linguistiques dans les pays unilingues et multilingues
   b) problèmes de langues difficilement accessibles

II. Moyens de réduction de la barrière linguistique

1. Cover-to-cover translations
2. Utilisation des traductions existantes
   a) collecte des traductions
   b) création de BD pour les traductions existantes
   c) problèmes de dépistage et de récupération des traductions existantes
3. Traduction
   a) traductions traditionnelles
   b) traduction de texte intégral par ordinateur
   c) traduction assistée par ordinateur

III. Institutions et traducteurs pour aider à réduire la barrière linguistique

1. Centres d'information et de documentation
   a) rôle de dispatcher
   b) coopération à la production de BD
   c) aide terminologique
2. Le traducteur et les problèmes de traduction
   a) compréhension de textes de haute technicité
   b) problèmes terminologiques
   c) dictionnaires et glossaires

IV. Le traducteur et la traduction assistée par ordinateur

1. Modernisation des méthodes actuelles de traduction
   a) traitement de textes
   b) imprimantes à alphabets multiples
   c) logiciels d'aide à la traduction
2. BD terminologiques
   a) Eurodicautom
   b) Termium
   c) glossaires et dictionnaires automatiques
3. Utilisation des BD bibliographiques en tant que source de terminologie actualisée
   a) BD multilingues; utilisation directe des thésaurus
   b) BD bibliographiques avec sélection par sujet et par langue

V. Conclusions

1. Solution des problèmes linguistiques du consommateur d'information
2. Information et aide au traducteur

MOTS-CLES

BD bibliographiques, BD pour les traductions, BD terminologiques, Cover-to-cover translations, Dictionnaires et glossaires lisibles en machine, Eurodicautom, Imprimante à alphabets multiples, Langue difficilement accessible, Logiciel d'aide à la traduction, Microprocesseur, Terminologie scientifique et technique, Termium, Thesaurus, Traduction assistée par ordinateur, Traduction par ordinateur, Traduction traditionnelle, Traitement de texte.

LANGUAGE BARRIERS : TRANSLATION PROBLEMS.

I. Language barriers and modern scientific and technical information

1. Problems to provide full-text documents and to get information

a) acces to documents
b) communication with owners

2. Problems of understanding litterature

a) problems of language understanding in unilingual and multillingual countries
b) problems with difficult accessible languages

II. Means for language barrier reduction

1. Cover-to-cover translations

2. Use of existing separate translations

a) collection of existing translations
b) DB production for existing translations
c) problems to find and to get translations

3. Translation

a) human translation
b) machine translation
c) computer-assisted translation

III. Organisations and translators providing help to reduce language barriers

1. Information and documentation centres

a) dispatching targets
b) cooperation for translation DB production
c) terminological help

2. Translator problems

a) comprehention of scientific and technical texts
b) terminological problems
c) dictionnaries and glossaries

IV. Translator and computer-assisted translation

1. Modernisation of translation processes

a) word processors
b) multiple-fonts printers
c) softwares for translation help

2. Terminological DB

   a) Eurodicautom
   b) Termium
   c) computerised dictionnaries and glossaries

3. Use of bibliographical DB as up-to-date terminological sources

   a) multilingual DB, direct use of thesauries
   b) use of bibliographical DB with selection by subject and by language

V. Conclusions

1. Solution of information users linguistic problems

2. Information and help to translators

KEYWORDS

Bibliographical DB

Computer-assisted translation

Cover-to-cover translation

Difficult accessible language

Eurodicautom

Human translation

Machine-readable dictionnaries and glossaries

Machine translation

Micro-computer technology

Multiple-fonts printer

Scientific and technical terminology

Software for translation help

Terminological DB

Termium

Thesauries

Translation DB

Word processor

Dans la littérature scientifique et technique, surtout dans les domaines de haute technologie, la terminologie est mouvante. La recherche, le secteur industriel et le secteur commercial sont confrontés à de nouvelles découvertes, de nouveaux produits, donc à de nouvelles appellations. Le manque de standardisation terminologique et l'utilisation de jargons industriels rendent la compréhension encore plus difficile non seulement lorsqu'il s'agit de langues étrangères, mais aussi pour les usagers pratiquant la langue du document.

La science de l'information, créée au profit des spécialistes, tend à couvrir exhaustivement la littérature scientifique et technique disponible dans le monde. La production de bases et banques de données informatisées est devenue pratique courante et en fait la seule solution d'information face à la masse croissante de documents.

On distingue deux genre de bases et banques de données : celles qui sont d'utilité universelle (CHEMABS, INSPEC, METADEX, NTIS, etc.) et celles qui sont élaborées pour un usage plus restraint, pour une industrie spécifique, quelques entreprises ou même une seule institution. Mais quelles qu'elles soient, leurs producteurs et utilisateurs se heurtent au problèmes de langue et de terminologie.

La standardisation terminologique elle-même, aussi bien que la production des bases de données, est étroitement liée à la communication entre les producteurs de documents, à la disponibilité de ces documents, au signalement de ceux-ci et à l'organisation d'une information cohérente.

La disponibilité des documents est problématique pour des raisons bien connues :

- concurence industrielle, sécurité, raison d'Etat;
- difficulté de dépistage des producteurs et détenteurs de documents et de contact avec ceux-ci.

La compilation d'une base de données constitue un premier pas vers une communication meilleure et vers une standardisation terminologique; en effet, le producteur de la base de données contacte les auteurs des documents, rassemble ces documents ou leur signalement, les indexe d'une façon uniforme, créant ainsi une terminologie plus ou moins homogène. En retour, les utilisateurs des bases de données se familiarisent avec une seule terminologie et ont tendance à l'utiliser.

L'utilisation universelle des bases de données par des usagers de différents pays et de langues différentes les habitue tous à une seule langue et une seule terminologie (en majorité l'anglais) et ouvre déjà une brèche dans le mur de l'incompréhension linguistique.

Il convient ici de souligner que les problèmes linguistiques sont différents d'un pays à l'autre : il y a des pays unilingues et multilingues.

Les langues aussi peuvent être divisées en deux catégories : les langues dites faciles : langues latines et germaniques, et les langues que l'on qualifie de difficilement accessibles : langues slaves, chinois, japonais, arabe, etc.); de ces dernières seules seront retenues celles qui véhiculent actuellement une production scientifique significative, comme par exemple le japonais et le russe.

Lorsque l'on diférencie les problèmes par pays on remarque que les pays unilingues tels que la France, la Grande Bretagne, les USA ont des problèmes avec toutes les langues; mais s'ils ont des difficultés de compréhension, en revanche ils ont depuis longtemps mis en place des structures et créé des organismes facilitant la traduction pour les langues des deux groupes. Les chiffres sont significatifs : 29% de toutes les traductions réalisées concernent les paires de langues français-anglais et allemand-anglais.

Les pays multilingues tels que la Suisse, le Canada, la Belgique ont beaucoup moins de problèmes de compréhension des textes originaux et ont de ce fait un réseau de traduction scientifique plus faible, axé surtout sur les langues difficilement accessibles.

Réduire la barrière linguistique ne signifie pas traduire systématiquement les textes dans la langue de l'utilisateur du moment. Lorsque l'on parle de réseau de traduction et de structure facilitant la traduction, peu nombreux sont ceux qui en réalisent l'ampleur et la diversité. Il est pourtant primordial que les utilisateurs d'information et les spécialistes de l'information les connaissent.

Il y a en premier lieu les traductions intégrales de documents (cover-to-cover translations) que l'on publie surtout pour les documents de langues difficilement accessibles (russe, japonais). Il existe aussi des ouvrages de référence qui les signalent.

Il y a également un nombre considérable de textes qui sont déjà traduits et peuvent être réutilisés ultérieurement plusieurs fois. Pour cela il faut les dépister, les collecter et les compiler sous forme d'ouvrages de références ou de banques de données pour les signaler aux usagers potentiels. La collecte de documents traduits se heurte aux difficultés déjà évoquées de concurrence, de sécurité.

C'est dans ces cas particuliers que doivent intervenir des groupements plus restraints garantissant un contrôle des utilisateurs, une défense des intérêts de chacun à l'intérieur du groupement ainsi que le secret professionnel pour les fournisseurs de traductions et leurs utilisateurs. Ce n'est que dans ce contexte, en respectant les facteurs évoqués qu'il est possible de rassembler les traductions déjà réalisées.

Il va de soi que la base de données est la solution la plus valable pour mettre ces traductions à la disposition des intéressés. Pour exemple il convient de citer le "World Translation Index", base de données qui propose les traductions rassemblées en Europe et aux USA par le National Translation Centre (USA) et l'International Translation Centre (Europe). Le WTI est également diffusé sous forme traditionnelle d'ouvrage de référence.

Certains groupements signalent pour leurs membres les traductions effectuées au sein du groupe, par exemple le Centre de Recherches Métallurgiques, Belgique.

En dernier recours pour abolir la barrière linguistique il y a évidemment la traduction. On distingue :

- la traduction traditionnelle;
- la traduction réalisée par ordinateur;
- la traduction assistée par ordinateur.

Le mode de traduction traditionnel fait appel aux traducteurs et présente certaines difficultés, surtout pour les textes techniques

très spécialisés : le traducteur se heurte à des problèmes de compréhension technique et de terminologie.

La traduction par ordinateur a fait de grands progrès, mais son efficacité et la qualité de la traduction sont encore discutables. De multiples études en ont parlé, évaluant ses avantages et ses défauts. On estime toutefois que ces traductions sont valables du moins pour l'information des scientifiques et à condition que le vocabulaire introduit dans l'ordinateur soit limité à un domaine bien déterminé. Il est utile de souligner le résultat positif obtenu pour l'indexdation des documents et la préparation des résumés à introduire dans les bases de données. Les textes standards ou répétitifs sont également assez bien traduits. L'avantage principal de la traduction par ordinateur est indiscutablement la possibilité de standardisation de la terminologie. Le phénomène de standardisation provoqué par les bases de données pour une langue se manifeste ici pour plusieurs langues. Malheureusement les reproches essentiels que l'on formule généralement sont considérables : prix trop élevé, perte de temps considérable en révision des textes, inaptitude de l'ordinateur à prendre en considération les nuances du texte.

La traduction assistée par ordinateur, quant à elle plus flexible, combine le travail créatif du traducteur et les possibilités de l'informatique. C'est pour cela qu'elle présente une solution avantageuse, valable, réalisable sans frais excessifs. Ses avantages sont évidents pour l'utilisateur des textes et pour le traducteur, le rendement et la qualité de la traduction s'en trouvent améliorés.

On constate que les possibilités de réduction de la barrière linguistique sont nombreuses; il sera donc utile de connaître les institutions qui les mettent en oeuvre.

Il y a avant tout les organismes de traduction et de diffusion des bases et banques de données :

- les producteurs de bases et banques de données;
- les centres d'information;
- les centres de signalement des traductions;
- les centres terminologiques.

Des organismes contactent les producteurs de documents, compilent la littérature, mettent l'information à la disposition de l'utilisateur, rassemblent les traductions existantes, produisent et exploitent des dictionnaires terminologiques informatisés. Ils sont utiles à l'usager puisqu'ils répondent de diverses façons à leurs besoins en information et abolissent la barrière linguistique. Ils sont également les auxilliaires du traducteur puisqu'ils sont impliqués dans la production et l'exploitation des sources terminologiques et des sources d'information sur les traductions existantes.

L'utilisateur d'information confronté aux proulèmes linguistiques et le traducteur qui doit l'aider ont tous deux le plus grand intérêt à les connaître et à recourir à leurs services.

Néanmoins lorsqu'il est question d'abolir la barrière linguistique c'est bien le traducteur qui est le personnage-clé. Ces problèmes méritent dès lors une étude détaillée.

Plus que jamais le traducteur se trouve confronté à des textes scientifiques très spécialisés et à des descriptions de haute technicité. Il est obligé de faire face aux problèmes

- de compréhension des textes;
- de recherche terminologique pour des techniques de pointe;
- de recherche des sources d'information;
- d'adaptation aux méthodes modernes de travail et de recherche de l'information.

La compréhension des textes est rendue difficile par la complexité des sujets qui pour 55% relèvent du domaine des sciences précises et de la haute technologie; la spécialisation très poussée des disciplines impliquées dans chaque sujet exigent une certaine formation scientifique ou tout au moins un contact étroit avec le spécialiste.

La recherche des termes équivalents dans la langue source et dans la langue cible est rendue extrêmement difficile à cause de la mouvance de la terminologie, du manque de standardisation dans les définitions de nouveaux produits et procédés, des jargons industriels : le traducteur consacre en moyenne 17% de son temps à la recherche d'information et 58% seulement à la traduction proprement dite.

Les dictionnaires - outils traditionnels du traducteur - présentent un décalage considérable par rapport aux textes proposés, surtout dans les domaines mentionnés et dans les disciplines de pointe.

L'essence même du dictionnaire et les méthodes d'édition ne permettent pas une mise à jour suffisamment rapide; on estime généralement qu'ils ont 4 ans de retard sur l'évolution des techniques.

L'adaptation aux nouvelles techniques est difficile, tributaire d'une information quasi-inexistante dans le monde de la traduction, d'apprentissages longs et d'investissements financiers consequents.

Il faut se rendre à l'évidence que malgrés ces réserves, seule la technique des ordinateurs peut à la longue résoudre tous ces problèmes. C'est à cette technique qu'ont eu recours les documentalistes et les spécialistes de l'information; c'est vers elle que doivent se tourner désormais les traducteurs. Pour cette raison il convient de les en informer le plus largement possible.

La traduction assistée par ordinateur est un concept qui reprend plusieurs aspects de l'informatisation et qui s'applique à plusieurs étapes du processus de traduction.

L'utilisation du PC (personal computer) permet au traducteur de :

- bénéficier de tous les avantages d'un traitement de texte;
- traiter le texte source et le texte cible en les visualisant simultanément sur l'écran;

- avoir des logiciels possédant plusieurs jeux de carractères (latin, cyrillique, grec, hébreu, etc.) et utiliser simultanément plusieurs d'entre eux, par exemple un texte en cyrillique avec dénominations latines pour un sujet médical;
- être raccordé par modem aux serveurs de son choix pour obtenir l'information terminologique et scientifique nécessaire;
- constituer ou acquérir des glossaires spécialisés;
- utiliser des logiciels d'aide à la traduction.

Actuellement 51% des traducteurs utilisent un équipement informatisé, mais la majorité l'utilisent presque uniquement pour le traitement de textes et les tâches administratives auxiliaires.

Les bases de données terminologiques, autre aspect de l'informatisation sont susceptibles de fournir une aide précieuse aux traducteurs. Malheureusement seuls 14% d'entre eux utilisent des bases de données terminologiques contre 91% qui continuent à se servir de dictionnaires spécialisés traditionnels.

Pourtant les banques de données terminologiques sont simplement des dictionnaires bilingues, parfois multilingues, parfois encyclopédiques, différents des dictionnaires traditionnels par le fait qu'ils emploient un support magnétique en lieu et place du support papier. Ils sont donc mémorisables et peuvent être manipulés par ordinateur. Ce sont des dictionnaires automatiques et ils possèdent tous les avantages des enregistrements pour ordinateurs : maniabilité, rapidité de réponse, facilité d'introduction de nouvelles données. C'est dire que ce genre de dictionnaire permet d'obtenir la réponse à une recherche terminologique en quelques secondes, mais aussi qu'ils peuvent être régulièrement complétés et mis à jour de façon permanente. Il est évident que ce nouveau type de dictionnaire sera très utile pour les traducteurs.

Actuellement il n'y a pas encore beaucoup de banques de données terminologiques; celles qui existent ne comportent pas encore assez de termes pour couvrir toutes les langues et tous les domaines des sciences, mais même telles qu'elles se présentent, elles ont déjà rendu bon nombre de services aux traducteurs.

Il convient de citer en exemple le volume de données des principales d'entres elles :

- TEAM, compilée par la firme Siemens, est une banque de données multilingue qui contient déjà plus de 500.000 termes;
- le dictionnaire multilingue encyclopédique EURODICAUTOM, produit par le service terminologique de la CEE, comporte près de 350.000 termes;
- la banque terminologique multidisciplinaire du gouvernement canadien TERMIUM est une banque de données bilingue français-anglais, anglais-français; son volume atteint les 800.000 termes.

Ces exemples montrent que si ces banques terminologiques ne sont pas encore exhaustives, le nombre de termes enregistrés les rend déjà exploitables avec un taux de réponses positives satisfaisant.

En tout cas leurs avantages sur les dictionnaires traditionnels sont incontestables : maniabilité, rapidité de réponse, mise à jour permanente. Deux facteurs importants militent aussi en leur faveur : la possibilité de dialogue en direct et la gratuité (dans plusieurs cas) due à une exploitation au stade expérimental et promotionnel.

On assiste aussi à une affluence sur le marché de glossaires plus spécialisés, plus petits, à choisir par le traducteur en fonction de ses besoins et à exploiter sur son propre PC.

Les bases de données bibliographiques représentent par définition une source terminologique inépuisable et actualisée; elles donnent souvent satisfaction là où les méthodes traditionnelles et les banques de données terminologiques n'ont pas fourni l'information recherchée. Ces bases de données répertorient la quasi-totalité des publications dans tous les domaines des connaissances : cela signifie que toutes les langues y sont représentées.
Il y a deux façons d'utiliser les bases de données pour des recherches terminologiques.

1. Lorsque la base de données elle-même est bilingue ou multilingue, il est possible d'obtenir une réponse directement. En effet, dans ce cas, les titres et les mots-clés y figurent dans toutes les langues utilisées. On peut citer comme exemple la base de données multidisciplinaire PASCAL qui est multilingue (français-anglais-allemand-espagnol). Il y a aussi la base de données TITUS (textile) qui est proposée en trois versions : française, anglaise et espagnole et qui comporte donc directement la terminologie de l'industrie du textile en trois langues. Cette première façon d'utiliser les bases de données permet d'obtenir les termes recherchés dans les langues de la traduction directement lors de la consultation.

2. La deuxième alternative concerne les langues qui ne sont pas utilisées lors de la production des bases de données : par exemple, le russe ou le japonais. Dans ce cas la solution consiste à sélectionner les ouvrages qui traitent le sujet soumis au traducteur dans la langue originale et dans la langue cible. A la différence d'un traducteur qui essaye de se documenter sans données de départ bien déterminées, l'ordinateur fournit immédiatement les références précises et pertinentes des ouvrages traitant le sujet présentant la terminologie recherchée.

Les avantages de cette approche par rapport aux méthodes traditionnelles qu'utilisent les traducteurs sont évidents : rapidité, pertinence des réponses et, en plus, possibilité de commander rapidement les ouvrages sélectionnés (par le même ordinateur) dans le cas où les bibliothèques les plus proches ne les possèderaient pas. Un avantage supplémentaire non négligeable : les ouvrages signalés par les bases de données sont récents, écrits par des spécialistes dans leur langue; cela permet au traducteur d'obtenir la terminologie usuelle, prise dans son contexte, au lieu de celle des dictionnaires, qui est parfois artificielle et inappropriée.

En plus par l'intermédiaire de divers centres d'information ou bureaux spécialisés les banques de données terminologiques et les bases de données bibliographiques sont accessibles aux intéressés même s'ils ne sont pas encore équipés en outils informatisés.

La comparaison des investissements en temps de recherche et en dictionnaires et des investissements éventuels en recherche par ordinateur est nettement favorable à la méthode nouvelle.

CONCLUSIONS

La réduction des difficultés linguistiques comporte deux aspects :
- la solution des problèmes du consommateur de l'information,
- l'information et l'assistance du traducteur.

L'abolition des barrières linguistiques est étroitement liée à l'organisation de l'information scientifique et technique et rencontre les mêmes problèmes : dépistage et compilation des documents, standardisation terminologique.

Les progrès de l'information informatisée diminue les barrières linguistiques en imposant une seule langue véhiculaire et une standardisation terminologique.

L'examen des problèmes linguistiques montre que l'approche du problème est différente pour chaque communauté linguistique et pour chaque groupe de langues.

Il est cependant évident qu'il faut mettre en oeuvre, signaler aux intéressés et utiliser tous les moyens visant à faciliter la compréhension : cover-to-cover translations, BD de traductions, etc. Il est tout aussi important d'atteler à cette tâche les producteurs et distributeurs d'information que les traducteurs.

Il apparaît que si les problèmes du traducteur sont complexes, les méthodes modernes, l'informatisation et surtout une bonne information des traducteurs sont susceptibles de les résoudre. La création d'outils nouveaux pour les traducteurs et l'utilisation rationelle de l'information mise à leur disposition est inévitable pour réactualiser le métier de la traduction.

La coordination de tous les efforts peut assurer la réalisation de l'ensemble des mesures qui permettrons une meilleure compréhension de l'information par les spécialistes, la création et l'implémentation d'outils appropriés pour faciliter le processus de traduction. Les barrières linguistiques s'en trouveront considérablement réduites.

REFERENCES

1. Smith, David J., The digital report on translation pratices, Language Monthly, N°42,march.1987, p. 11-13.
2. Kuzjmin, Ju. G., Voprosy terminologii v praktike i teorii perevoda nauchno-tekhnicheskoj literatury, Metodika informatsionnoj raboty, N°6.1969, p. 11-19.
3. Barré, Laurence, Compte-rendu de la Réunion d'information organisée par la SFT le 28 novembre 1986 sur les centres serveurs et banques de données, Traduire, N°131,mars.1987, p. 17-20.
4. Proceedings of the World Systran Conference Commission of the European Communities, Luxemburg, 11-14 february 1986.

LINGUISTIC BARRIERS: TRANSLATION PROBLEMS

I. CLEMENS,
Ir. civilian électr.,
Head of the Belgian Translation Centre,
CNDST-CBT,
4 Bld de l'Empereur
B-1000 Brussels (Belgium)

SUMMARY

Linguistic problems are in close connection with information problems. They therefore are examined jointly in order to solve the terminological problems and to reach the documentary mass. The problems of comprehension of the texts by the specialist are examined as well as the difficulties raised by the translator. Linguistic problems are examined according to the characteristics of the unilingual and multilingual countries and according to the groups of languages. The existing means to reduce linguistic barrier and implied institutions are considered from the point of view of the user and of the translator. The translator is the keyman to overcome linguistic barrier. He is confronted with the problems of scientific and technical text comprehension, of search for information and technical terms sources, of adaptation to the modern work methods. The use of data processing provides him some help and multiple possibilities of rationalization of the translation process. Among the known methods of translations, computer-assisted translation seems currently the most performing: use of a PC for translation process as such, use of computerized dictionaries and bibliographical data bases as multilingual terminology sources reliable and up-to-date. The whole of raised measures makes it possible to reduce linguistic difficulties of the information user and to ensure an effective help to the translator.

I. Linguistic barriers and modern scientific and technical information.

1. Problems to provide full-text documents and to get information
   a) access to documents;
   b) communication with the information owners.

2. Linguistic problems
   a) linguistic problems in unilingual and multilingual countries;
   b) problems with difficult accessible languages.

II. Means for language barrier reduction.

1. Cover-to-cover translation;

2. Use of the existing translations;
   a) collection of translations;
   b) DATA BASE production for existing translations;
   c) problems to find and to get translations.

3. Translation
   a) human translation;
   b) machine translation;
   c) computer-assisted translation.

III. Institutions and translators providing help to reduce the linguistic barrier.

1. Information and documentation centres.
   a) dispatcher targets;
   b) cooperation for DATA BASE production;
   c) terminological help.

2. Translator and problems of translation.
   a) text comprehension with high technicality;
   b) terminological problems;
   c) dictionaries and glossaries.

IV. Translator and computer-assisted translation.

1. Modernisation of the current translation methods.
   a) word processing;
   b) multiple-fonts printers;
   c) software for translation help.

2. Terminological DATA BASES.
   a) Eurodicautom;
   b) Termium;
   c) Computerised glossaries and dictionaries.

3. Use of bibliographical DATA BASES as up-to-date terminological sources.
   a) Multilingual DATA BASES; direct use of the thesaurus;
   b) Bibliographical DATA BASE with selection by subject and by language.

V. Conclusions.

1. Solution of the information users linguistic problems;.

2. Information and help to the translator.

KEYWORDS

Bibliographical DATA BASE,
Translation DATA BASE,
Terminological DATA BASE,
Cover-to-cover translations,
Machine-readable dictionaries and glossaries,
Eurodicautom,
Multiple-fonts printers,
Difficult accessible languages,
Software for translation help,
Microprocessor,
Scientific and technical terminology,
Termium,
Thesaurus,
Computer-assisted translation,
Machine translation,
Human translation,
Word processing.

In scientific and technical literature, especially in the fields of high technology, the terminology is moving. Research, the industrial sector and the commercial sector are confronted with a new discoveries, new products, therefore with new names. The lack of terminological standardization and the use of industrial jargons make comprehension even more difficult not only for foreign languages, but also for the users who practice the document language.

Information science, created to specialists, tends to cover exhaustively the scientific and technical literature available in the world. The production of bases and computerized data banks has become practical current and is in fact the only solution of information in relation to the increasing document mass.

Two kinds of bases and data banks are distinguished: those of universal utility (CHEMABS, INSPEC, METADEX, NTIS, etc) and those worked out for a more restricted use, for a specific industry, several firms or even only one institution. But whatever they are, producers and users come up against language and terminology problems.

Terminological standardization itself, as well as the production of data bases, is closely dependent on the communication between the document producers, the availability and announcement of these documents, and the organization of coherent information.

The availability of documents is problematic for well-known reasons:

— industrial competition, safety, reason of state;
— difficulty to find the documents' producers and owners and get in touch with them.

A data base compilation constitutes a first step towards a better communication and towards terminological standardization; indeed, the producer of a data base contacts the documents' writers, gathers the documents or their annoucement, indexes them in a uniform way, thus creating a more or less homogeneous terminology. In return, the users of the data bases get used to only one terminology and tend to use it. to use it.

The universal use of the data bases by users of various countries and of various languages makes all of them used to only one language and only one terminology (in majority English) and already breaks through the wall of linguistic incomprehension.

It is advisable here to stress that the linguistic problems are different from a country to another: there are unilingual and multilingual countries.

Languages also can be divided into two categories: easy said languages: Germanic and Latin languages, and languages that one describes as difficult accessible: Slavic, Chinese, Japanese, Arab languages, etc); of these last only will be retained those which convey currently a significant scientific production, such as for example Japanese and Russian.

When one differentiates the problems by country, one points out that the unilingual countries such as France, Great Britain, the USA have problems with all languages; but if they have difficulties of comprehension, on the other hand they have set up structures for a long time and created organisms facilitating the translation for languages of both groups. The figures are significant: 29% of all the translations carried out concern the pairs of French-English and German-English languages.

The multilingual countries such as Switzerland, Canada, Belgium have much less comprehension problem of the original texts and have consequently a weaker network of scientific translation, centred especially on difficult accessible languages.

Language barrier reduction does not mean to translate systematically the texts in the language of the user. When one speaks about network translation and structure facilitating translation, very few realize its width and diversity. However it is essential that information's users and specialists know them.

There are initially the full-translations of documents (cover-to-cover translations) that one publishes especially for the difficult accessible languages (Russian, Japanese). They are also announced by reference works.

There is also a considerable number of texts which are already translated and can be re-used several times at a later date. For that, one should detect them, collect them and compile them in the form of reference works or data banks to announce them to potential users. The translated document collection comes up against the already raised difficulties of competition and safety.

In these particular cases have to intervene more restricted groupings to guarantee control of the users, defence of the interests of each one inside the grouping as well as professional secrecy for the translation suppliers and their users. It is only in this context, by respecting the raised factors that it is possible to gather the translations already carried out.

It goes without saying that the data base is the most valid solution to put these translations at the disposal of the interested parties. For example it is advisable to quote "World Translation Index", data base which proposes the translations gathered in Europe and in the USA by the National Translation Centre (USA) and the International Translation Centre (Europe). WTI is published also in the form of a traditional reference work.

Certain groupings point out for their members the translations carried out within the group, for example the Metallurgical Research Centre, Belgium.

As a last resort to overcome linguistic barrier, there is obviously the translation. One distinguishes:
- human translation.
- machine translation;
- computer-assisted translation.

The traditional translation methods requires translators and raises certain difficulties, especially for the very specialized technical texts: the translator encounters problems of technical comprehension and terminology.

Computer-translation made major progress, but its effectiveness and quality of translation is still debatable. Multiple studies spoke about it, evaluating its advantages and its defects. It is considered however that these translations are valid at least for scientists' information and provided that the vocabulary input into the computer is limited to a properly determined field. It is useful to stress the positive result obtained for documents indexingation and preparation of summaries to be introduced in data bases. Standard or repetitive texts are also rather well translated. The principal advantage of computer-translation is indisputably the possibility of terminological standardization. Standardisation phenomenon caused by data bases for one language appears here for several languages. Unfortunately the essential reproaches that are generally formulated are: price too high, considerable waste of time in post-edition, computer inadpitude to take into consideration text nuances.

Computer-assisted translation, as for it more flexible, combines translator's creative work and data processing possibilities. That is why it presents an advantageous, valid, realizable solution without excessive expenses. Its advantages are obvious for texts' user and translator; translation output and quality are improved.

It is noted that the possibilities of linguistic barrier reduction are numerous; it will be therefore useful to know the institutions which implement them.

There are above all translation and diffusion organisms of the bases and data banks:

- base producers and data banks;
- information centres;
- translation announcement centres;
- terminological centres.

Organisms contact the document producers, compile literature, put information at user's disposal, gather the existing translations, produce and exploit computerized terminological dictionaries. They are useful to the user since they answer in various ways their information requirements and overcome linguistic barrier. They are also the translator's auxiliaries since they are implied in the production and exploitation of the terminological and information sources on the existing translations.

The information user confronted with linguistic problems and the translator who has to help him have both a major interest to know them and to resort to their services.

Nevertheless, when it is question of overcoming barrier linguistic, the translator is actually the keyman. These problems deserve consequently a detailed study.

More than ever the translator finds himself confronted with very specialized scientific texts and with high technicality descriptions. He is compelled to cope with the following problems:
- comprehension of the texts;
- terminological research for advanced techniques;
- research of information sources;
- adaptation to modern work methods and information retrieval.

Texts comprehension is made difficult by the complexity of subjects which for 55 % are relevant to precise sciences field and high technology; the very high specialization of the disciplines implied in each subject requires a certain scientific training or at least a close contact with the specialist.

Research of the equivalent terms in source and target language is made extremely difficult because of terminology changes, lack of standardization in defining new products and methods, industrial jargons: the translator devotes on average 17 % of his time in information retrieval and 58 % only to the translation as such.

Dictionaries — the traditional translator's tools — present a considerable shift in relation to the proposed texts, especially in the fields mentioned and in advanced disciplines.

The very purpose of a dictionary and edition methods does not allow any sufficiently rapid update; it is generally considered that they are 4 years late on technological developments.

The adaptation to the new techniques is difficult, dependent on almost not-existent information in the world of translation, long trainings and consequent financial investments.

One should admit that in spite of these reserves, only computer technique can at last solve all these problems. It is to this technique that had resorted documentalists and information specialists; translators have to consider it now. For this reason it is advisable to inform them the most widely possible.

Computer-assisted translation is a concept which takes again several aspects of computerization and which applies at several stages of the translation process.

The use of the PC (personal computer) allows the translator:

- to benefit from all the advantages of text processing;
- to process the source and target text in visualizing them simultaneously on the screen;
- to have software having several character sets (Latin, Cyrillic, Greek, Hebrew, etc) and to use simultaneously several of them, for example a text in Cyrillic with Latin denominations for a medical subject;
- to be connected by host modem of his choice to obtain necessay terminological and scientific information;
- to constitute or acquire specialized glossaries;
- to use software for translation help.

Currently 51 % of the translators use computerized equipment, but a majority of them uses it quite only for text processing and auxiliary administrative tasks.

Terminological data bases which are another aspect of computerization are likely to provide precious help for translators. Unfortunately, only 14 % of transaltors use terminological data bases against 91 % who continue making use of traditional specialized dictionaries.

However, terminological data banks are simply dictionaries bilingual, sometimes multilingual, sometimes encyclopaedic, different from the traditional dictionaries by the fact that they employ a magnetic medium instead of paper medium. They are therefore memorized and can be handled by computer. They are automatic dictionaries and they have all the advantages of the recordings for computers: answer handiness, speed, input facility of new data. It is to tell that this kind of dictionary makes it possible to obtain the response to terminological research in a few seconds, but also that they can regularly be supplemented and updated in a permanent way. It is obvious that this new type of dictionary will be very useful for the translators.

Currently there are not yet many terminological data banks; those which exist do not yet comprise enough terms to cover all languages and all the fields of sciences, but even as they are, they have been already very useful to the translators.

It is advisable to mention data size of the most important data banks:

— TEAM, compiled by the firm Siemens, is a multilingual data bank which contains already more than 500,000 terms;
— the multilingual encyclopaedic dictionary EURODICAUTOM, produced by the terminological service of the EEC comprises nearly 350,000 terms;
— the multidisciplinary terminology data bank of the Canadian government TERMIUM is a bilingual French-English, English-French data bank; its volume reaches the 800,000 terms.

These examples show that if these terminological data banks are not yet exhaustive, the number of recorded terms makes them workable already with a satisfying positive answer rate.

In any case their advantages over traditional dictionaries are undeniable: answer handiness, speed, permanent update. Two important factors militate also in their favour: possibility of on-line dialogue and free of charge (in several cases) due to an exploitation at an experimental and promotional stage.
One attends also to a multitude of more specialized and smaller glossaries on the market, to be chosen by the translator according to his needs and to to be exploited on its own PC.

The bibliographical data bases represent by definition an inexhaustible and update terminological source; they often give satisfaction where traditional methods and terminological data banks did not provide required information. These data bases index almost all the publications in all the fields of knowledge: that means that all languages are represented there. There are two ways of using the data bases for terminological research.

1. When the data base itself is bilingual or multilingual, it is possible to obtain an answer directly. Indeed, in this case, the titles and the key words appear in it in all languages used. One can quote as an example the multidisciplinary data base PASCAL, which is multilingual (French-English-German-Spanish). There is also the data base TITUS (textile) proposed in three versions: French, English and Spanish and containing therefore directly the terminology of textile industry in the three languages. This first way of using data bases makes it possible to obtain the required terms in languages of the translation directly during consultation.

2. The second alternative concerns languages which are not used during data bases production: for example, Russian or Japanese. In this case the solution consists in selecting the documents which cover the subject submitted to the translator in source and target language. Contrary to the translator who tries to document himself without well defined starting data, the computer provides immediately the precise and relevant references of the documents covering the subject presenting the required terminology.

The advantages of this approach in relation to the translator's traditional methods are obvious: speed, relevance of the answers and, in addition, possibility of ordering quickly the documents selected (by the same computer) if the closest libraries did not have them. Considerable additional advantage: the documents announced by the data bases are recent, written by specialists in their language; that enables the translator to obtain usual terminology, taken in its context, instead of that of the dictionaries, which is sometimes artificial and inappropriate.

Moreover, via various information centres or specialized offices, terminological data banks and bibliographical data bases are even accessible to the interested parties if they are not equipped yet in computerized tools.

The comparison between investments in research time and dictionaries and possible investments in computer research is definitely favourable to the new method.

CONCLUSIONS

Linguistic difficulties reduction has two aspects:
— solution of the information user problems,
— translator's information and assistance.

Linguistic barriers reduction is closely dependent on the organization of scientific and technical information and encounters the same problems: search and compilation of the documents, terminological standardization.

Progress of computerized information reduces linguistic barriers by imposing only one common language and terminological standardization.

Linguistic problems study shows that the approach of the problem is different for each linguistic community and for each group of languages.

It is meanwhile obvious that one should implement, announce to the interested parties and use all the means aiming to facilitate comprehension: cover-to-cover translation, DATA BASE translations, etc. It is as important that information producers and providers as translators undertake this task.

It appears that if the translator's problems are complex, the modern methods, translators' computerization and especially good information are likely to solve them. The creation of new tools for the translators and the rational use of the information placed at their disposal is inevitable to bring up to date translation profession.

Coordination of all the efforts can ensure to achieve the whole measures which will allow better comprehension of information by specialists, creation and implementation of tools adapted to facilitate the translation process. Linguistic barriers will be considerably reduced.

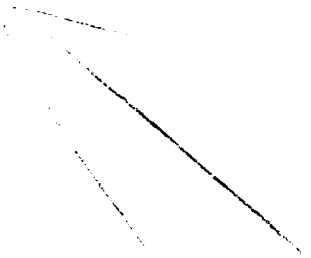
REFERENCES

1. Smith, David J., The digital report on translation practices, Language Monthly, N° 42, march 1987, p.11-13.

2. Kuzjmin, Ju. G., Voprosy terminologii v praktike i teorii perevoda nauchno-tekhnicheskk literatury, Metodika informatsionnoj raboty, N 6, 1969, p. 11-19.

3. Barred, Laurence, report of meeting organized by SFT on 8 November 1986 on host centres and data banks, Traduire, N 131, March 1987, p. 17-20.

4. Proceedings of the World Systran Conference Commission of the European Communities, Luxemburg, 11-14 February 1986.



Technical Change needs Organizational Change

by
David Boddy
University of Glasgow
Glasgow Business School
53-59 Southpark Avenue
GLASGOW G12 8LF

INTRODUCTION

In this paper, I will try to show that effective performance cannot be assured simply by installing expensive, state of the art technology. Our studies of many applications of computing and information technology have included both successful and unsuccessful examples (1). The clear practical lesson from this work is that the successful cases have usually been those there technical change has been accompanied by appropriate organizational change. The less successful ones have generally been those where projects have been dominated by technical considerations, with little or no thought given to organizational ones.

These lessons seem as relevant to defence installations as to the civilian ones in which our research has been conducted. The common characteristic is that all organizations are confronted by technical developments, which are intended to help them meet their objectives. For them to succeed in that task, it appears to us that management needs to ensure that three big hurdles are successfully crossed. These are:

Managing the project

Setting the right objectives

Changing the organization

Each of these hurdles will be outlined in turn, and the paper concludes with some proposals on how they can be overcome.

MANAGING THE PROJECT

Who's in charge?

A safe assumption is that the direction a project takes will reflect the interests and preferences of the function or group to whom top management allocate project leadership. A good example of this occurred when a manufacturing company set out to re-organize its mainframe computing system, involving a move from a heavily centralized system to one in which computing services were provided at each of their several facilities.

The old system had, as is usually the case, been run by the finance function, and, without a great deal of thought, the Board decreed that the project to establish the new centralized facility would also be run by the accountants. The new system worked adequately: but production and other line managers argued that they had been able to make very little contribution to the design of the system. The effect was that in designing the system major opportunities were missed to include facilities which would have been of great value to production staff at no extra cost.

A similar debate can occur over whether the project is led by a computer expert or a user. In general, successful applications have been those where the users have been in a dominant position in the project teams, with, of course, competent computer expertise available. A sure way to get a system that doesn't meet the practical requirements of a function is to give too much authority in shaping the project to the technical experts.

Finally, great care needs to be taken to ensure that project teams as a whole are not only balanced in terms of their representation from key areas, but also that they possess adequate teamworking skills. We know of one project which failed dismally because it was entrusted to a project team which was not only functionally imbalanced, but which lacked the skills necessary to resolve problems, take decisions, and get action.

Who do they consult?

As a general rule it is worth assuming (until it is proved otherwise) that the staff most closely involved in operating existing procedures and systems have the best understanding of them, and of how a more advanced computer system would help. More senior staff, at a distance, tend not to be aware of, or to have forgotten, the practical operating realities of daily work. If they take too dominant a role in designing and selecting new systems, there is a severe risk that they will computerize a myth, rather than a real operation.

A successful example of involving staff in specifying a new information system occurred in a food company. They had previously installed a traditional management information system, producing huge quantities of print-out, which no-one used. A new senior management decided to abandon this system, and

to start again. They set up a series of temporary teams in all major areas of the factory consisting of both operators and staff. Everyone was given the opportunity to think through their job, and to indicate what information would help them to perform more effectively. This basic data was worked into a coherent specification, which system staff then used as their target. The final system, radically different from the one it replaced, was successfully implemented, and continues to be a daily help to staff at all levels.

A similar approach can be taken when the decision concerns the evaluation and purchase of standard systems. Within obvious budgetary or compatibility constraints, experienced staff who will be using the equipment will probably make a better choice than either their bosses or the technical experts, and not to include them in the decisions on what to to buy is to ask for trouble.

How to develop competence?

Contrary to popular belief, sophisticated equipment generally needs skilled, trained and committed staff, if it is to be used effectively. Whilst most staff will be able to cope with the routine aspects of a new system with little or no training, many of the potential benefits are lost if deliberate steps are not taken to develop, and to keep developing competence.

We have often found companies which, having spent very large sums on hardware and software, regard training as a cost to be minimised or avoided altogether if possible. This is a false economy, which many have acknowledged in their discussions with us: asked what lessons they have learned from a project, the one most frequently cited is that they would spend more time and money on training. Appropriate funds need to be built into the overall capital budget, to ensure that acquiring competence is seen as being just as important as acquiring technology.

It is also important to ensure that people are given time to develop skills. If they, or their managers, are put under constant pressure for output above all else, training time will inevitably be diminished. Similarly, reliance on other operators, also under production pressure, to do the training rarely works. Such people can certainly help the training and familiarisation process: but they will experience unreasonable stress if they are expected to train other staff properly and meet their own output targets. Finally, it is important to consider what new competencies are need by those less directly involved with the new system, as well as those who are operating it.

SETTING THE RIGHT OBJECTIVES

An operating or strategic emphasis?

Whatever the technical characteristics of a new application may be, it is management which establishes what kind of contribution it will make to the business. In some cases, the emphasis has clearly been on using the technology to improve operating performance in the production and delivery of current goods and services. The benefits sought have been those which could be measured, and where quantifiable improvements over current practice could be confidently expected. Thus new technology is used if it promises to reduce labour costs, save energy, overcome bottlenecks in production, reduce scrap and so on. In the same vein, computer technology has often been seen as a way of increasing the amount of control which can be exercised over operations, by allowing procedures to be specified more precisely, to reduce dependence on scarce skilled labour, or to ensure a more even and regular pace of work.

Elsewhere, the objectives have been different. Technology has been introduced to secure less tangible benefits, such as improving the quality and timeliness of information available to managers, with the intention of enabling more confident decisions to be made. In other cases technical changes have been used to offer new or significantly enhanced products or services, and thus radically shift the position of the organization in the eyes of its customers or clients.

We have distinguished these two approaches as having either an "operating" or a "strategic" emphasis (2, 3, 4). Which approach is taken is of vital significance to the way the capabilities of IT are used in the organization, but is, by and large, independent of the technology. The key factor is senior managers' awareness of the scope for strategic uses of emerging technology, and its willingness to embark on this much more difficult approach.

Adopting a strategic approach is a much greater challenge than that of using technology primarily to improve current operations. It presents those responsible for a project with new conceptual and organizational challenges. For example, managers need to have a global and consistent picture of the enterprise, and of how technical developments can be implemented which will radically change the kind of tasks it is able to do, and the services it is able to provide.

Exploiting this potential also requires managers who are able to conceive the operation of the organization as a whole, unconstrained by established departmental or other structures. Most managers are familiar with well-established lines of authority, and vertical information flows. Technology makes it possible to move information independently of established structures - and an ability to see and value this possibility is essential if the strategic capabilities are to be achieved.

Does the system fit the objectives?

Even if senior staff are aware of the strategic possibilities, there is a danger that those in charge of procurement and installation unintentionally subvert those objectives. One company with which we worked planned, at senior level, to introduce an elaborate, highly automated manufacturing system, with a very clear set of strategic objectives in mind (quality, rapid response to design changes, ability to meet small orders quickly and so on). Unfortunately, these objectives were not fully appreciated down the line, where a series of decisions was taken about the details of the hardware, software and operating layout, which clearly reflected an emphasis on cutting costs rather than adding value. They ended up with an expensive system which, in practice, provided few benefits over that which it replaced: and certainly was not consistent with the strategic vision which had prevailed at the start of the project.

Does the Organization encourage strategic action?

Senior management has a major role to play in creating a climate within the organisation that encourages innovation in support of strategic objectives. Staff at middle levels need to be willing to make the effort and take the risks associated with technical change. Acting as the promoter or champion of a major technical change, especially if it needs to be accompanied by organizational change, puts the person concerned in a highly vulnerable position. They need to master not only the technical aspect, but also be willing to argue the case through the organization, often in the face of indifference or outright opposition.

Why should they bother? In many organisations, rewards go to those who play to the rules, keeping existing practices running smoothly, and who don't stir things up. Such organizations are unlikely to encourage people to take vulnerable, risky positions: to that extent, they are unlikely to cross the hurdle of establishing coherent strategic objectives for their technological advances.

CHANGING THE ORGANIZATION

The more scope there is for computer technology to make possible big changes in the services offered, the more likely it is that there will need to be changes in the shape and style of the organization. Decisions need to be taken on these matters, as well as on those of a technical nature. What are the options, and what experience is there elsewhere?

Structures and Relationships

Technical change makes it possible to eliminate some functions, alter the importance of others, and to create new ones - it becomes possible, and perhaps essential, to shake up existing boundaries of responsibility, and existing relationships. For example, advanced manufacturing systems are greatly enhancing the role of support facilities like planning, tooling and maintenance. Because of the much greater dependence of line managers on these facilities, two companies we have worked with are changing structures and reporting relationships to make these traditionally separate support systems closer to, and more answerable to, line management.

Technical change also makes it possible to link together successive stages of a process more closely, with data being transmitted automatically around the organization. There is a strategically important choice here: because a thing is technically possible does not mean it has to be implemented. For example, a hospital laboratory could have linked its automated analytical equipment directly to a newly installed computer system which stored information on patients, processed the results, and produced the reports on individual samples. Such a link would have made the later stages of the process somewhat more convenient: but it was rejected as the analytical process would then have been completely dependent on the computer being up and running before work could start. This was judged on unacceptable risk, and a decision was taken to include a deliberate "gap" in the system which would be filled by an operator, rather than introduce a fully automated procedure.

Management style

Choices also arise over how the management style of an organization is changed as technology is introduced. For example, computer technology can be used either to centralize decision-making, or to decentralize it. The temptation is often to take the former option, those in positions of power generally taking the view that the more influence and control they have, the better the organization will work.

The fallacy of this presumption is well illustrated by the example of a decision taken by the Board of a multi-national company not to increase the centre's knowledge of the periphery. They were building a new headquarters building, in which they would incorporate a significant amount of office automation. All the suppliers whom they invited to submit proposals try to persuade the company to install systems which, as well as automating routine head office functions, would enable them to have direct computer links to all of their overseas subsidiaries. This would have enabled head office to receive much more frequent reports on matters like sales, or operating costs in their remote plants, thus, it was argued, improving management control.

The Board decisively rejected such notions. The success of the business, they felt, had been based on giving substantial autonomy to local managers. They were accountable to the company in broad terms, and there was, of course, close monitoring of their quarterly and annual figures. But within those constraints they were left to run the business. Putting in direct computer links to head office would undermine that autonomy, and was therefore not implemented when the rest of the office automation system was installed.

Work Organization

Having skilled committed people depends on ensuring that their jobs are organized in a way that complements rather than replaces human skill and experience. Many of the less successful applications of computer technology have been those where work has been organized in a way that has led to boredom, inattention and loss of interest amongst the staff concerned. Such attitudes cannot produce high performance.

Conversely, where decisions on work organization have ensured that the skills and experience of staff have been complemented by the technology, the effects have been highly beneficial. In such cases, staff have shown themselves able and willing to accept responsibility, to work to overcome temporary difficulties in the system, to suggest valuable additional uses, and so on. By organizing work in a way that satisfies well established motivational requirements of individual jobs, and grouping them together into natural work teams, companies have been able to achieve far more benefits from technical innovation than where such factors have been ignored.

GUIDELINES FOR INNOVATION

In order to overcome these hurdles which lie in the way of successful innovation, the following guidelines are proposed. They are based on successful practices in a wide range of organizations, and can serve as a checklist for those charged with introducing technical change.

Nine guidelines for exploiting the innovative potential
of new computing technologies and information systems

PURPOSE:

1. Ensure that the use of new technology has a clear **strategic focus** which targets long term market objectives and not just current internal operating problems.

2. Review and implement **positive policies** in areas like employment practice and investment appraisal, so that they encourage innovation and the strategical use of new technology.

3. When evaluating hardware and software options, make sure that you choose **kit to fit** the current and anticipated needs of the business.

PEOPLE:

4. Review **work organization** in relation to new technology, to encourage flexibility, creativity and skills development.

5. Review **management styles** and working arrangements, to ensure that these are consistent with the strategic goals of the business, and with achieving the potential benefits of new technology.

6. Design **support systems** which are consistent with strategic aims, and which enable support staff to contribute in flexible and creative ways to the changing needs of the organization.

PROCESS:

7. Establish a clear **project management** responsibility to guide the often protracted implementation process.

8. Plan the nature and timing of user **involvement**, to ensure that key staff establish "ownership" of new equipment and procedures.

9. Develop a systematic training plan to equip users at all levels with the **competence**, in the form of new skills and knowledge, they need to exploit the innovative potential of the new technology.

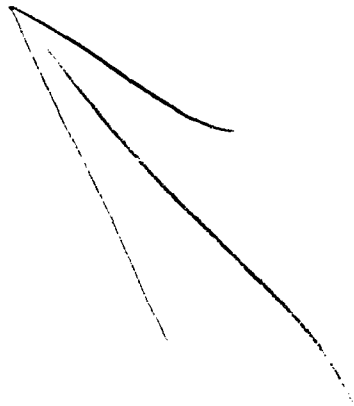
REFERENCES

(1) Boddy, D. and Buchanan, D.A., The Technical Change Audit, Manpower Services Commission (1987).

(2) Benjamin, R.I., Rockart, J.F., Scott Morton, M.S. and Wyman J.H., "Information Technology. A Strategic Opportunity", Sloan Management Review Spring, pp 3-10 (1984).

(3) King, W.R., "Developing Strategic Business Advantage from Informational Technology" in Piercy, N. (Ed.) Management Information Systems: The Technology Challenge, London, Croom Helm (1986).

(4) Porter, M.E., and Miller, V.E. "How information gives you competitive advantage", Harvard Business Review, 63, July-August, pp 149-160 (1985).



USING STANDARDS TO BREAK DOWN INFORMATION TRANSFER BARRIERS

Cynthia J. Durance
Director
National Library of Canada
395 Wellington St.
Ottawa, Ontario
K1A 0N4
CANADA

Summary

Standards for representing and encoding data as well as standards for the interconnection of systems having different hardware and software are both essential for effective electronic information transfer. To date, although major strides have been made internationally on standards in both the data and telecommunications areas, much remains to be done before effective electronic information transfer can take place either nationally or internationally. This paper addresses the difficulties in achieving approved standards internationally, the state of the art, and some new standards that potentially will have a positive impact on information transfer such as those pertaining to electronic publishing. It also addresses the Open Systems Interconnection Reference Model, a suite of telecommunications standards which hold great promise for the interconnection of systems with different hardware and software. Finally, the paper discusses the work that has been done to date to implement these standards for library and information applications and challenges for the future.

Standards have been discussed either implicitly or explicitly by previous speakers as an important element in reducing barriers to information delivery. Therefore, my paper is intended as a overview of the groups of standards which are pertinent to electronic information transfer and the current state of these standards and their implementation.

Standards which are essential to achieving effective electronic information transfer fall into two broad categories: technological standards and standards for the encoding of the data which is stored, retrieved and transmitted by the technology.

During the past five years, innovations in technology have made us increasingly aware of the possibilities and the need for provision of effective on-line as well as off-line transmission of data to remote sites.

In 1977, the Comité consultatif international télégraphique et téléphonique (CCITT) began work on a suite of international standards designed to provide, not only interconnection, but also intercommunication between hardware and software of different types and makes. This suite of standards is under the umbrella of the Open Systems Interconnection (OSI) Reference Model. Very early in the process the CCITT and the International Organization for Standardization (ISO) jointly endorsed the OSI Reference Model: the first time in history there had been such a joint endorsement of a standard between these two organizations.

There are several important characteristics about this suite of standards which make it potentially very powerful. First, the Reference Model is a 7 layer model where each layer is dependent on the layer below. (see Figure 1). These standards specifications must be implemented within software on each machine that needs to communicate with other machines. The standards are designed as communications standards, therefore they logically work between the front-end of one computer and another, and are invoked only when computer-to-computer communication is required. In other words within your own system you are free to manipulate data as you wish, as long as that data can be translated into the standard data definitions and communications format when required. Therefore, local autonomy is preserved: an important ingredient to the success of any standard.

Secondly, the development of the OSI standards has from the beginning been driven and supported by the telecommunications and the hardware and software industry. This is important to us as potential users of the standards because it means that software implementations of them will be available "off the shelf" from the industry. Moreover, industry members have created organizations to ensure that implementations of these standards across the industry can, in fact, interwork. The Corporation of Open Systems (COS) in North America, the Standard and Promotion and Application Group (SPAG) in Europe, and Promotion of Open Systems Interconnection (POSI) in Japan have all been created within the past five years in order to achieve this goal. An international umbrella organization for these regional organizations is currently being set up.

Since the OSI suite of standards is a very complex and hierarchical one, we, as users, could not afford the investment necessary to implement all seven layers of them if the substantial development costs were not able to be amortized in large part by the computer industry.

However, this is not to say that as potential users of OSI standards we can simply sit back and wait for them to be available. At the application (7th) layer the standards for the software are application specific and therefore need to be developed within specific user industries. These application level standards or protocols define the messages to be sent between computers, the format of the messages and their sequence for each application. We in the information sector need to be concerned with the development of OSI protocols for book order, for interlibrary lending, for document delivery, etc. and need to have input ino the development of more generic application protocols such as File Transfer and Management (FTAM), Information Search and Transfer, Distributed Directories protocols etc.

The National Library of Canada has since 1979, had a research and development team working on the development of OSI library and information application layer protocols. To date, we have developed and implemented a file transfer protocol to transfer Marc records between an IBM and CYBER, GEAC and Honeywell machines each running different software and operating on systems at the NLC and 3 Canadian universities. This software has been in operation since November 1984 on the four machines. An interlibrary loan messaging protocol has also been developed and implemented at the NLC to allow tne automatic sending and receipt of interlibrary loan messages as well as the tracking of them. This application layer protocol, which is now in the Canadian, American and international standards process, has been implemented using electronic mail for the lower six layers of the protocol.

In addition work is underway on a book order protocol, a cataloguing application and a series of protocols for distributed directory systems. The National Library is currently working with GEAC Computer Corporation to implement the protocols on the new software they are designing.

During this same period the Linked Systems Project (LSP) in the U.S.A. was formed consisting of the Library of Congress, the On-line Computer Library Centre (OCLC) Inc., Research Libraries Group (RLG), and the Western Library Network (WLN). They have developed an Information Search and Retrieval protocol and a record transfer protocol which have been in operation since 1986.

While there has been constant liaison between the work at the NLC and that of the LSP, the approach has been somewhat different.

In the U.S. they cho e to build themselves layers four to six of the protocols, relying on the telecommunications carriers for layers one to three only, while in Canada we chose to use what was available from the telecommunications industry for all of layers 1 to 6, secure in the knowledge that our national carriers were committed to implementing OSI. NLC will convert to the pure OSI protocols for these lower layers in concert with the telecommunications companies. The other major difference is that all LSP protocols have been developed using the connection-oriented mode, while in Canada, with the exception of file transfer, all the protocols have been developed using the store and forward (electronic mail) mode of connection (see Figure 2). Both modes are OSI compatible and as time goes on these differences will be resolved. Already the agreement between NLC and LC to work toward common protocol standards for library applications is beginning to bear fruit and it is anticipated that shortly the Canadian Interlibrary Loan protocol and the U.S. Information Search and Retrieval protocol will both be North American standards.

On a more general level, the wide implementations of the OSI electronic mail standard (CCITT Recommendation X.400) which is already underway will lead to the interconnection of electronic mail systems nationally and internationally. This development will provide possibilities for new and improved delivery of information.

In my opinion, OSI will be the communications technology of the next decade for on-line communication, bringing with it a very substantial reduction in the current barriers to information transfer.

However, it is essential that we understand that the ability to connect systems alone will not achieve useful results - in addition, the data content and format transmitted must be in standard form for intercommunication to take place. In this regard, the library and information community are fortunate in having already achieved a high degree of concensus on data and format standards internationally. This achievement positions the community well to take advantage of OSI technology. Sectors which do not yet have widely recognized standards for data will find their ability to take advantage of this technology for their applications circumscribed.

In the last fifteen years we have achieved, and have widely implemented, several standards for encoding bibliographic data in machine-readable form. Spearheaded by the library profession there are the International Standard Bibliographic Descriptions (ISBD's) for nearly all types of materials and these are used globally. These de facto standards prescribe the recording of title page information but not how to provide access points to a bibliographic record. They also differentiate the data elements in a record by the use of punctuation symbols which are particularily useful for display of data recorded in foreign languages.

Further, the International Standard Book Number (ISBN) and the International Standard Serial Number (ISSN) are both ISO standards and are widely implemented and used by libraries, publishers and distributors, and by the abstracting and indexing community. In the U.S.A., the ISSN is now required for registering serials to qualify for reduced postal rates, and the ISSN will also be used on bar codes and article number identification within serials in the USA, UK and elsewhere. The ISBN is already widely in use in barcoding schemes.

The MARC format for machine-readable encoding, while imperfect, has been widely accepted as the communications format for the tape exchange of bibliographic data amongst libraries. It is based on an international standard-ISO 2709.

Designed originally by the Library of Congress, other national versions of this format emerged: CANMARC, UKMARC, INTERMARC (France), AUSMARC and MABEIN (West Germany) are examples. Recently there has been agreement internationally on a superset format, UNIMARC, which is increasingly being used for international tape exchange.

The abstracting and indexing community commonly use the UNISIST Reference Manual for recording contents of articles in serials, and the development by UNESCO of the Common Command Format (CCF) has provided a communications format for this data.

While these standards greatly facilitate a reasonably effective method of communicating bibliographic data nationally and internationally for machine-readable tape transfer, there are still problems. The UNISIST Reference Manual and the CCF are not totally compatible with UNIMARC, and perhaps never will be because their underlying purpose is different. Although the ISBN, because of its simple structure (one edition of a book, one number), has been able to be implemented simply by assigning blocks of unique numbers to publishers who then assign them to their publications, the ISSN has been more difficult to manage. Because serials, by their nature, change title, merge, split, etc. the ISSN must be bibliographically controlled by national and ultimately an international agency who maintain an identification record for serials. This is the International Serials Data System, a UNESCO program hosted by the French government with currently 45 national centres throughout the world feeding the International Centre information made accessible via the ISDS Register of Serial Publications. While this register of, currently, 370,000 records is a tribute to international cooperation, problems exist mainly due to delays in assigning and making accessible the assignment of ISSN's and key-titles. Work is underway to alleviate these problems, but until timeliness is accomplished, the use of ISSN as a global identification number for serials will not be as widespread as is needed.

In spite of, or because of, the increased activity in international data definitions and format standards in several sectors of the information industry coupled with the OSI standards activity, there will be an increased need to convert data from one standard to another especially when communicating across various sectors of the information industry. Current methods of format conversion are time-consuming and slow and must be improved if we are to be able to exploit the potential of on-line data transfer.

Hence work has been done in the government of Canada to use an artificial intelligence technique called inferencing to write software to improve the format conversion process. The prototype has been successfully developed but more work is needed before such developments can aid the conversion of machine readable formats.

Equally the need for a means of standardizing search arguments and keys for searching different data bases has long been recognized as a desirable tool for easier use of multiple data bases. Various attempts such as the creation of a common command language (CCL) or mediation software, have met with some resistance from information providers who felt its implementation would reduce the specificity and clarity of their search strategy. Recently, a Standard Query Language (SQL) has been approved internationally which allows sophisticated search arguments to be programmed, but is not very "user-friendly". SQL and CCL may be used in conjunction with each other, CCL being the display mechanism.

In addition to format converter software being potentially a very useful tool for conversion of record formats, it is possible that, in future such software could also be employed to convert a search query into formats appropriate for searching data bases employing differing search strategies and thus largely obviate the need for techniques such as a common command language. If this supposition proves correct and the process proves economical, the impact of such software on improved service could be substantial. At present, it is not possible to predict whether either or both of the above methods will prove efficient or effective.

Further, major strides are being made in the field of data definition for electronic publishing. For the sake of discussion it is useful to divide electronic publishing into two processes, electronic data creation, and electronic dissemination. Greater progress has been made to date in the former than the later.

The acceptance by ISO of the Standardized General Markup Language (SGML) in 1986 has provided a standard for encoding data for word processing packages, thereby allowing discs to be read on more than one make of word processing software during the editing and publishing process. Further the Association of American Publishers (AAP) in the U.S. has developed a standard for electronic manuscripts which is an application of the SGML standard and which will provide a common means of describing front matter and back matter, (i.e. title page table of contents, and reference information) for an electronic manuscript. This standard, which will be submitted to ISO shortly, will be of great importance to those of us concerned with control of and access to publishing in electronic form because it will provide a standard means of providing bibliographic control for electronic publications. It is called the Electronic Manuscript Preparation and Markup standard. It would appear that CDROM may be a more widespread method of publishing electronically than on-line access, although both will undoubtedly exist.

CDROM and optical disc technology is a technological development which is complementary to the OSI developments. Optical disc technology has great potential for the dissemination of large quantities of information without the need to rely on telecommunications services, and with the associated cost savings, particularly internationally. While agreement has been reached on standards for imprinting the discs, such agreement is still lacking for the software to retrieve information from the disc. Until a retrieval standard is in wide use, the ability to search discs from multiple vendors will be cumbersome and costly, and will detract from the use of optical disc technology for information transfer.

In summary, increased activity in both technological and data standards development world-wide mirrors a recognition that without both data and telecommunications standards it will be impossible to develop effective and efficient mechanisms for the interchange of data. The barrier of lack of standards to effect information exchange is being slowly overcome by the dedicated work of experts from many countries. It is incumbent upon us, the users, to be involved in this process to ensure the standards meet our needs and also to implement them faithfully in order to provide our clientele with ready and timely access to the information they require, regardless of the data source.

FIGURE 1

OPEN SYSTEMS INTERCONNECTION (OSI)
REFERENCE MODEL

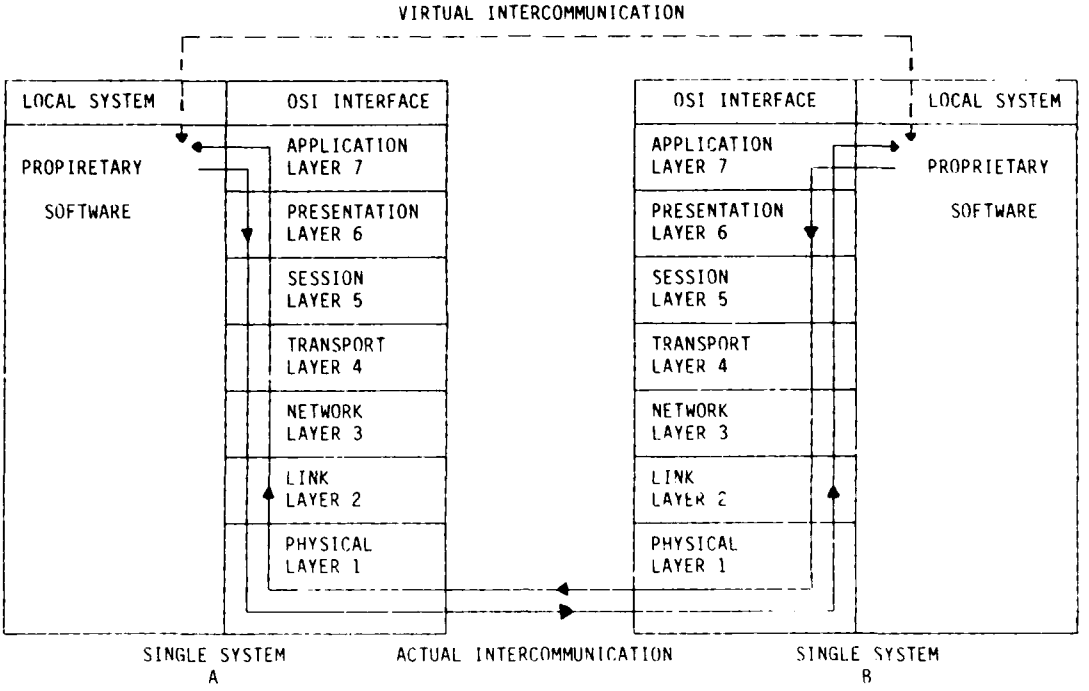
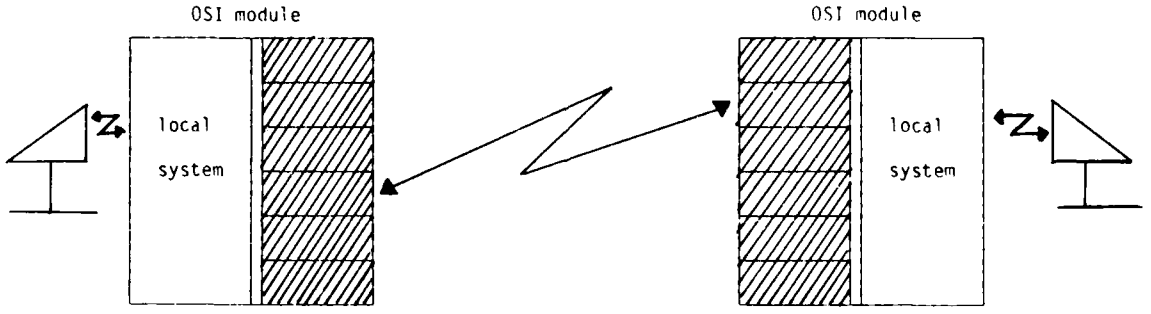
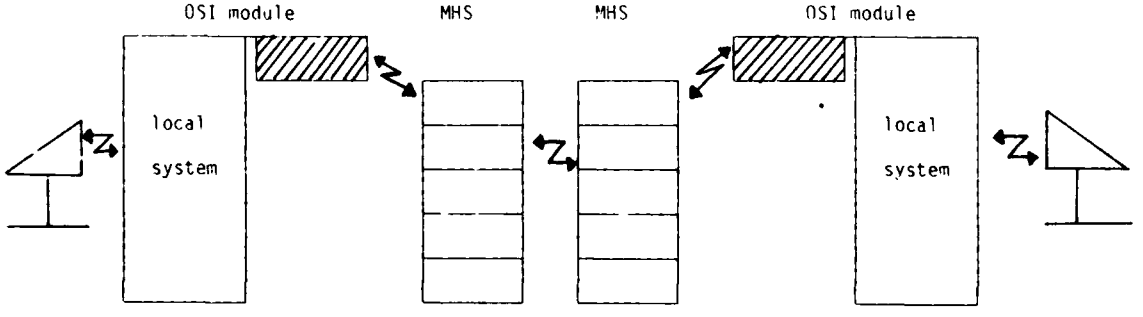


FIGURE 2

TWO MODES OF INTERCONNECTION



a) Connection-oriented - single session per transaction.



b) Connectionless or message-oriented - separate sessions to complete a transaction.

NOTES ON THE TRANSLATIONS

The Information and Documentation Department of Aérospatiale at Suresnes near Paris is currently experimenting with machine translation, using SYSTRAN. Since one of the main barriers considered during this meeting was the barrier of language, they kindly offered to translate into English the two papers that were presented in French, and into French one of the English language papers. For the latter, we selected the paper by Professor Zimmerman (No.4), because it was concerned, to a large extent, with machine translation. AGARD is extremely grateful to Aérospatiale for the time and trouble they have taken over this task.

This section of the Conference Proceedings contains comments by Aérospatiale (in French and in English) and examples of the raw, unedited, output from SYSTRAN. It must be stressed firstly that SYSTRAN is only one machine translation system — there are a number of others — and secondly that the examples given were translated without the making of a special dictionary for the subject matter. When reading the translations it must also be borne in mind that the 'raw' translation output by the computer is intended merely to give an end-user (scientist, engineer, manager, etc.) a translation that is sufficiently good for him to grasp the meaning of the original document.

The times shown in the table on page T-3 give an indication only of the results of a single experiment. They should not be considered as being generally applicable to all translations. The comparison figures given in the last column of the table are estimates of how long a professional translator might take to produce a raw quality translation from the same original text.

* * *

COMMENTAIRES SUR LES TRADUCTIONS PRESENTEES

La Division d'Information et de Documentation d'AEROSPATIALE à Suresnes procède actuellement à des expériences de traduction machine à l'aide du système SYSTRAN. Etant donné que l'une des principales barrières à la communication étudiée lors de la Conférence a été la barrière de la langue, AEROSPATIALE a proposé de faire traduire vers l'anglais les deux communications présentées en français et de faire traduire l'une des communications présentée en anglais vers le français, à l'aide du système SYSTRAN. En ce qui concerne cette dernière proposition nous avons choisi la communication du Professeur ZIMMERMAN (No.4) puisqu'elle concerne principalement la traduction machine. L'AGARD est très reconnaissant envers l'AEROSPATIALE du temps passé à l'accomplissement de cette tâche, et de la peine qu'elle s'est donnée pour la mener à bien.

Cette partie du compte-rendu de conférence comporte des commentaires émis par AEROSPATIALE (en français et en anglais) et quelques exemples de traductions SYSTRAN brutes et sans post-édition. Il faut d'abord souligner que SYSTRAN n'est qu'un système de traduction machine parmi d'autres et que les exemples donnés ici ont été traduits sans la compilation préalable d'un dictionnaire spécialisé. Lors de la lecture des traductions il doit être tenu compte du fait que la traduction "brute" de l'ordinateur est destiné à un utilisateur (scientifique, ingénieur, gérant etc.) qui n'a besoin que de comprendre le sens général du document original.

Les temps passés à faire les traductions, indiqués au tableau ci-après, ne représentent que les résultats d'une seule expérimentation. Ils ne sont point applicable à toutes les traductions. Les chiffres indiqués à la dernière colonne représentent des estimations du temps que passerait un traducteur professionnel à faire une traduction "brute" de qualité égale à partir du même texte original.

COMMENTAIRES GENERAUX

par

Oleg Lavroff
Société Aérospatiale
Chef du Département Information — Documentation
B.P. No.76
92152 Suresnes Cedex
France

L'expérimentation menée à l'AEROSPATIALE a pour but de pouvoir diffuser à terme le plus rapidement possible et à moindre coût la traduction de textes présentant un intérêt immédiat et contenant des informations nécessaires à la bonne marche de l'enterprise.

L'expérimentation est menée avec le système SYSTRAN. Les textes joints montrent le niveau actuel de développement de ce système et mettent en évidence les points particuliers qui doivent être développés (élargissement des dictionnaires actuels utilisés par SYSTRAN) et améliorés (amélioration du système à partir des analyses effectuées par les traducteurs).

Au stade actuel, les traductions brutes ne sont pas suffisamment compréhensibles et elles nécessitent de ce fait une post-édition (vérification et mise en forme minimale correcte des phrases) de l'ordre de 15 minutes par page pour atteindre un niveau suffisamment compréhensible. Ce temps devrait être réduite de moitié une fois résolu le problème de la terminologie (codification des termes) et des principaux défauts constatés dans le système, comme indiqués ci-dessous.

A ce jour, la traduction issue de la post-édition correspond au niveau souhaité de traduction brute fournie par la machine. Cette post-édition est assurée par des traducteurs qui participent volontairement à l'expérimentation de la TAO (Traduction Assistée par Ordinateur) à l'AEROSPATIALE.

Le rôle actuel du traducteur (correcteur) consiste:

(1) à vérifier la traduction brute machine et à la rendre suffisamment compréhensible,

(2) à analyser les textes bruts traduits (signaler le manque de terminologie et les erreurs de traduction dues à la machine).

A terme, une fois le système mis au point, on sera amené à considérer deux types de traducteurs:

(1) un correcteur/post-éditeur (vérification et correction des traductions brutes machines), qui aura pour mission de vérifier la traduction de textes devant être diffusés très rapidement et pour lesquels une traduction de qualité humaine n'est pas nécessaire (suivi de la veille technologique, la concurrence etc....).

(2) un traducteur qualifié réservé à des traductions présentant des difficultés, que seul un traducteur humain peut résoudre.

Le tableau ci-après indique les temps passés pour la saisie des texte.., lorsqu'ils ne sont pas disponibles sur support informatique (reconnaissance des caractères plus corrections), le temps de post-édition et le temps global de la traduction (hors transfert et temps machine qui sont difficilement chiffrables et peu significatifs).

Pour information on indique également le temps que passerait un traducteur humain (traduction de qualité humaine brute) afin de pouvoir estimer à quel temps se situe une traduction brute machine suffisamment compréhensible.

Dans l'hypothèse la plus favorable à l'avenir (sans post-édition, qualité brute compréhensible):

— temps de traduction par la machine serait d'environ 15—20% par rapport à une traduction qualité humaine brute.

Réf. Textes	Nbre de Mots	Lect Optique	Temps Correct	Temps Post-éd	Temps Global	Temps Trad. qual humaine brute
La barriere linguistique (I. Clemens) (Fr. → angl.)	3395 mots (13 pages de 250 mots)	10 mn	30 mn	7 h 30	4 h 10	4 h 00
Linguistic-technical aspects of machine translation (H. Zimmermann) (Angl. → franç.)	4969 mots (20 pages de 250 mots)	10 mn	30 mn	6 h 30	7 h 10	6 h 00
Les barrieres linguistiques et culturelles au transfert de l'information (Roland Larue) (Fr. → angl.)	4136 mots (16 pages de 250 mots)	13 mn	40 mn	7 h 00	7 h 53	4 h 48
	12.500 mots	33 mn	100 mn	17 h 00	19 h 13	14 h 48

GENERAL COMMENTS

The purpose of the experiment being carried out at AEROSPATIALE is to achieve the fastest and cheapest possible translation of documents of immediate interest, containing information of a type which is necessary for good company management.

The experiment is being carried out using the SYSTRAN system. The attached documents show the present level of development of this system and highlight those points which require further attention (expansion of the current SYSTRAN system dictionaries) and improvement (refinement of the system by means of analyses made by the translators).

The raw translations currently being produced are not sufficiently understandable and as a result, require post editing (checking and minimal sentence restructuring), in order to reach an acceptably intelligible level, which requires something of the order of 15 min per page. This time should be halved once the problem of terminology (coding of terms) has been overcome and the main faults found with the system, as indicated in the following pages, have been rectified.

At the present time, the translations produced after post-editing represent the quality level aimed at for future raw machine translation. This post-editing is carried out by translators who have chosen to take part in the CAT (Computer Assisted Translation) experiment at AEROSPATIALE.

The translator's (corrector's) present rôle is to:

(2) Check the raw machine translation and make it sufficiently understandable

(2) Analyse the raw translations (pointing out any gaps in terminology and the errors in translation made by the machine)

Eventually, once the system has been perfected, we shall be able to talk about two types of translator:

(1) A corrector/post editor (checking and correction of raw machine translations), who will check documents which need to be issued rapidly and for which human translator quality is not required (monitoring of state of the art in the relevant field, activity of competitors etc.)

(2) A qualified translator reserved for translations containing difficulties which only a human translator can resolve.

The table below shows the time spent on document input, these texts not being available on data processing media (ie time for character recognition plus corrections), the post-editing time, and the global time taken for the translation (exclusive of transfer times and machine times which are difficult to cost and are not very significant.

For comparison we also show the time which would be spent by a human translator to produce the same quality of translation.

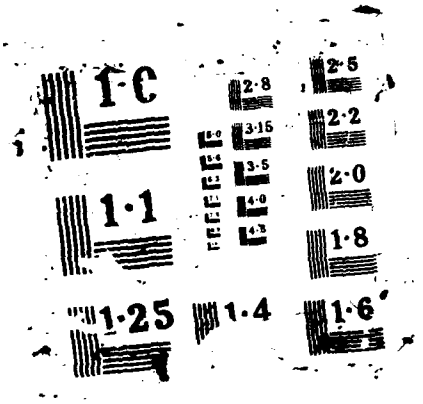
Assuming the most favourable conditions in the future (no post-editing, comprehensible raw machine translation), the translation time by computer would be about 15—20% of the time for human raw quality translation.

Text	No of Words	Optical reading	Correcting	Post-editing	Global	Human raw quality translation
La barrière linguistique (J. Clemens) (Fr. → Eng.)	3395 words (13 pages of 250 words)	10 mn	30 mn	2 h 30	4 h 10	4 h 00
Linguistic-technical aspects of machine translation (H. Zimmermann) (Eng. → Fr.)	4969 words (20 pages of 250 words)	10 mn	30 mn	6 h 30	7 h 10	6 h 00
Les barrieres linguistiques et culturelles au transfert de l'information (Roland Larue) (Fr. → Eng.)	4136 words (16 pages of 250 words)	13 mn	40 mn	7 h 00	7 h 53	4 h 48
	12 500 mots	33 mn	100 mn	17 h 00	19 h 13	14 h 48

AD-A194 653 BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION CO..(U) ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT NEUILLY.. MAR 88 AGARD-CP-430 2/2

UNCLASSIFIED F/G 5/2 NL

END
DATE FILMED
8 88


1·C
1·1
1·25
2·8
3·15
3·5
1·4
2·5
2·2
2·0
1·8
1·6

EXTRACT FROM UNEDITED MACHINE TRANSLATION OF PAPER 2.

Since the appearance of the language, the inhabitants of the sphere communicated always between them by local languages, or even of the dialects, which were really comprehensible only by a relatively low number of persons setting up a cultural regional group.

The twentieth century saw the conjugation of several phenomena: explosion of demography, lengthening of the life expectancy, increase in the trade and development of technology, which led the five thousand million men living on ground in 1987, to be communicated between them on several levels.

At the local level, language — or dialect — is always the communication tool of the cultural group of an area, with meanwhile of the new characteristics: the number of inhabitants of certain areas increased much during the twentieth century (Algeria, Egypt, Indonesia, Pakistan...) and the frequent bijectivity between cultural language and a state led to a strengthening of membership of this cultural group.

At the overall level, trade in any kind which develops since a few tens years on the whole of the world was, in point of fact, precedeed during two or three important centuries of trade but limited to what one could call them "large areas" sphere. And as it well necessary was to comprehend oneself to trade or beat oneself, one notes that today eleven languages are, in fact, become "vehicular" for one of these areas:

Mandarin	China
Indi	India
Englishmen	Northern America and Western Europe
Spaniard	America south and centre
Arab	Northern Africa and the Middle East
Malais	Asia of the south-east
Russian	Eastern Europe
Frenchmen	Western Europe and Africa
Swahili	East Africa
Japanese	Japan
Portuguese	South America

1.2 Tendency

The whole of the major demographic, socio-economic and military balances was always determining for the importance of language, and so is it difficult to forecast of the future. Meanwhile, if, for the next century, the assumption of major cataclysm is excluded (biological, military, geological or cosmic), the demographic factor becomes leading and done to appear profound changes in the importance of the major languages of the sphere.

Then the most outstanding facts for the fifty years to come are:
— the stabilization of certain languages: Mandarin, indi, French, Russian, Japanese,
— the growth of swahili, malais and Portuguese,
— the rapid growth of Spanish and the even more rapid growth of Arabic who could exceed Mandarin in the middle of the next century.

To some extent share of currently very important languages in number, like Mandarin, the indi and English will decrease to the profit primarily of Spanish and especially of Arabic.

In addition, political strategic choices, like the decision, in 1928, of Mustapha Kemal ATATURK to impose the Roman alphabet in Turkey and the decision to rock on the Cyrillic alphabet for esquimaux of U. r. S. S, are at the same time particularly important in the plan of exemplarity and of low importance at the quantitative level of the evolution of languages for the whole of the sphere.

2 THE USE OF LANGUAGES IN 1987 IN THE FIELD OF SCIENCES AND TECHNIQUES.

2.1 Need for the international scientific paper:

For a country, the broad cultural aims of scientific and technical information were synthetized in a circular (6) of the Minister for French research in September 1981:

— to express the scientifico-culturelle identity of people,
— to ensure the access of this people to modernity, in its own language,
— to ensure democratically the equality of the access to knowledge by the passage of the scientific and technical information on the whole of the social fabric of a nation.

This third point not being only the recovery of a concept of philosophical ALAIN which developed the idea that "culture level of a country was not defined by that of Plato but well by that of its being useful".

For a research organism, this international communication is also necessary for the defence of the quality of the organism (its image) or more simply for the justification of the use of the allocated funds. We know all that the major international periodic congresses are in fact the occasion given to the research organisms to give an account of their work in front of public opinion. At the level of the individual finally, the publication remains always necessary, despite the reality of the "invisible college" made up by the persons who are concerned with a same research orientation, for the anteriority problem arises often, as, quite simply, the consolidation of the personal the research worker's prestige; already several tens years ago only the dilemma publish gold perish is well anchored in the head of all the scientists.

In any event, it is well in science and in technology that intellectual, technological and financial investments are the largest ones and therefore, overall, it is well there that the need for best information is the most pressing one. In addition, a report of the D. O. E . (7) on the impact of scientific and technical information show that investment in I S. T. is particularly profitable for research.

2.2 Type of the communication: oral, written:

All the analyses (1, 2, 3, 4, 5, 6) reveal that English is, in fact, the most practical common language for the majority of the scientists.

It is currently the most widespread language quantitatively among the scientists and technicians of the whole world, and it is, additionally, located in the group of languages which, in the field of the quality of the publications are the most interesting ones.

English being therefore currently the common language of the scientific paper, this brings a fundamental difference between the English language scientist and the others. With regard to the oral communication (3, 4), and particularly during the meetings, congresses or conferences, anglophones have an enormous natural advantage which makes some non anglophone large losers of the communication on all the levels. At the level of of course the message, but more especially with that of the emotional relation: the loss, would say MAC LUHAN, is double, at the same time on the "message" and on the "massage".

This phenomenon is still amplified by the fact that many English or American scientists comprise only English and that they impose to some extent, to the whole of a group, the choice of the English language where they are of course the more at ease. We have all in memory situations where a group discussing in German, French or Spanish rocks on English on arrival of only one person, to the major disadvantage of all the autres*.

With regard to the written communication (3, 4), the situation are reversed and turn to the advantage of non anglophone.

Indeed (3), 12 % of anglophones can read French and 4 % German, whereas 97 % of the French-speaking persons and of German-speaking can read English. This enables theretore the non anglophone research workers to have access to the anglophone literature which represents 50 % production of the publications.

This result is corroborated by the analysis of "science index quotation" which shows that the anglophone authors quote with 90 % of the anglophone documents (whereas the production is of 50 %) and the possible difference in quality cannot explain a so major difference, more especially as two phenomena come into play in opposite direction: the fact that much the major international reviews are in English language is compensated by the fact that the documents bought by the libraries in language other than English much better is selected (owing to the difficulty of access) therefore, in doubt, better quality, and therefore more important to read.

A rather remarkable point of the written communication which does not appear me to have been pointed out enough is that certain specialists in the reading committee of a major international review are informed of the contents of Article six month before the other readers. All occurs as if information went to the best informed.

2.3 Type of contents:

Inside sciences and techniques, the linguistic situation are not the same one of a field to the other.

If we take the example of the French scientists (3), we note:

— that in the fields of mathematics, of physics and of chemistry, more than half of French author articles are written in English and that this percentage made only grow for the last years;

— that in the fields of the "sciences applied" (biology, sciences of the ground, medicine and technology), article to three is written in English, and that the growth of the number of the publications in English language seems to have stopped.

In addition, we note that the share of Japanese language grows quickly, particularly in technology. A report of 1986 (8) considers at 600 000 the number of scientifico-technologiques publications a year, of which less than 3 000 are the subject of a translation, even partial.

These two examples permit the following official report: the publications in fundamental sciences are directed towards writing in English , those of applied sciences are primarily written in national language.

* I remember a negotiation meeting in Brussels assembling 9 persons of 6 countries: 8 originating in 5 countries speaking usually French and American speaking only English.

EXTRACT FROM UNEDITED MACHINE TRANSLATION OF PAPER 4.

Considérant le développement technique dans l'industrie de l'information, on doit réaliser que l'ordinateur sera intégré dans presque chaque activité humaine professionnelle. Dans le domaine du mot ou du traitement de textes, l'interaction homme-machine, sur le niveau de PC ou sur la base d'un poste de travail sur le terminal au lieu d' une machine à écrire, sera la norme dans un proche avenir, qui signifie de la fin des années '80.

De l'autre côté, le problème de la traduction automatique de texte naturel en général — même pour ne pas parler sur la compréhension de la parole et la traduction de la parole — ne sera pas résolu dans le sens de HQFT (la traduction automatique entièrement automatique de haute qualité), en raison de la complexité des langages naturels. Donc il y a des limites (plus ou moins dans l'approche scientifique) et des possibilités des outils techniques d'intégration en cours de (l'humain) traduction ou même texte comprenant par les appareils.

Le but de cette contribution est d'explorer les possibilités de l' utilisation de l'ordinateur dans le domaine de l'appareil et de la traduction automatisée. Même si il pourrait être d'une certaine attraction traiter également de principaux problèmes de la translatabilité les textes (et de la compréhension), nous nous concentrerons sur la traduction technique, qui signifie la traduction de textes techniques ou communs.

Aspects systématiques

Même si il pourrait être d'un certain intérêt comment la traduction automatique est réalisée, le jeu linguistique d'aspects (particulièrement les modèles de la grammaire et des stratégies possibles ou employées) une partie subordonnée. Tellement on peut supposer qu'il y a une sorte de "boîte noire", où un texte ou des mots dans un langage naturel est mis dedans et, avec ou sans l' interaction humaine un ou les traductions alternatives sortent du système. La traduction pourrait être bonne, utile ou mauvaise sous les aspects de l'utilisateur.

Se concentrant sur l'utilisation et l'utilité de la traduction automatique ou de la traduction assistée par ordinateur, on doit — en principe — distinguer entre deux groupes d'utilisateurs principaux:

(1) le soi-disant utilisateur final, g de E.. un essai habile d'obtenir les informations d'un article écrit dans un langage naturel plus ou moins inconnu à lui; une personne écrivant une lettre à un ami en langue étrangère... et

(2) un agent professionnel, particulièrement un traducteur humain qui essaye d'utiliser l'appareil comme outil en cours d'accomplir son travail.

Sous ces aspects, on peut laisser de côté des projets et de base au sujet de la recherche selon les principes traduction automatique/ compréhension de la langue et orienter sur les outils pratiques resp. systèmes.

Comme point de départ, il est important de distinguer entre deux stratégies principales: traduction automatique (MT) et traduction assistée par ordinateur (CHAT).

— Un système s'appellera un système de traduction automatique (MT) seulement si un processus de traduction — le démarrage du texte source lisible sur ordinateur est accompli sans aucune interaction humaine pour atteindre une qualité de texte cible au moins "assez bonne" pour l'information

— Un système s'appellera un système de traduction assistée par ordinateur (le CHAT) si une interaction humaine est nécessaire ou prévue pour atteindre le but d'une "bonne" traduction d'un texte source (lisible sur ordinateur ou pas).

Il est bien évident qu'un système MT peut être employé, en plus, comme composant d'un CHAT : un texte peut être ajusté avant que les commencements MT pour obtenir de meilleures traductions automatiques ("pré-éditant"), et /ou un texte traduit par ordinateur peut être post-édité par les traducteurs humains pour obtenir la qualité plus élevée.

Il y a beaucoup de systèmes sur MT et du CHAT sur le marché affirmant d'être l'outil pratique droit, et le choix est difficile à être fait sans avoir des critères précis pour la décision.

Puisque la quantité de données dans les dictionnaires automatisés — sur la longue ligne de développement des systèmes MT et de CHAT — est le composant décisif, la mise à jour du dictionnaire (électronique) joue un rôle substantiel dans les deux solutions de rechange:

— Un concept peut être — particulièrement si le système a toujours une grande capacité d'entrées de dictionnaire que l'utilisateur, g de E.. le traducteur ne devrait pas participer à un tel processus (voir, par exemple le Systran-concept), de sorte qu'il y ait des spécialistes requis du côté d'administration de système pour faire le travail d' améliorer la base de données de dictionnaire;

— L'autre manière est la laisser à l'utilisateur professionnel achever le dictionnaire du système ou ajouter un vocabulaire spécial (voir, comme exemple, le logo-concept).

Il y a d'autres aspects de système qui jouent un rôle important sur le processus de décision: la disponibilité des couples de langues, la possibilité de en utilisant ou manipuler des types de texte spéciaux (g de E.. minutes, lettres,...).

Critères d'évaluation

Qualité

Même si une normalement boîte (ou doit) manipuler le système comme "boîte noire", là est des différences dans la qualité des résultats (crus) "purs" de traduction d'appareil (la production). Il n'est pas très facile de donner une mesure précise, mais il y a certains critères importants à noter (voir, pour les détails, en particulier. le concept Van SLYPE (1982) et les descriptions des deux évaluations SYSTRAN, Van SLYPE (1979)).

Les critères principaux sous ces aspects sont:

(1) fiabilité et fidélité, ces moyennes: à quel degré (mauvais, bon assez, bon) la teneur/signification du texte original est conservée..

(2) compréhensibilité et l'intellegibilité, ces moyennes: à quel degré l'utilisateur final peut lire et comprendre le texte traduit.

Ceci regarde assez simple, mais les problèmes se produisent dans le détail: Ainsi, sur la syntaxe et le niveau stylistique, un système normalement produira les traductions qui sont pires que les traductions humaines, tandis que sur le niveau lexical, particulièrement en identification du bon term (technique s), la traduction d'un système peut même être plus précise et consistant qu'une traduction humaine.

Environnement d'application

MT ou un CHAT doit être vu dans l'environnement (technique) des possibilités d'application. Ainsi, normalement une décision est non seulement la qualité orientée, mais également basée sur les possibilités de l'intégration dans un texte ou un système complexe de traitement de textes. Sous ces aspects, les composants suivants doivent être considérés:

Intégration dans les bases de données (bibliographiques ou textuelles)

Entretemps, les bases de données sont — techniquement parlé — accessibles world-wide par l'intermediaire des réseaux et même satelite de la communication de commutation par paquets. Ainsi, surmontant les barrières linguistiques, g de E.. entre l'anglais et le Japonais, mais particulièrement sur le marché multilingue européen, devient un désir important. Des efforts expérimentaux pour intégrer des systèmes MT dans un tel processus de l'information sont faits au Japon: comme un exemple la base de données inspec qui est originellement anglaise peut être interrogée avec les mots clés japonais. Les mots clés et plus tard (anglais) le titre sont automatiquement traduits pendant le dialogue en Japonais (Nagao et autres. 1982). D'une manière semblable, un système MT pour l'allemand à l'anglais est employé par l'intermédiaire d'un processus en lots et avec la post-édition) pour traduire les titres des bases de données allemandes (Zimmermann et autres 1987).

Il est bien évident que la traduction de titres et d'abrégés pourrait être également faite par les traducteurs humains. Mais il y a certains arguments pour MT et CHAT: le texte pour être traduit est lisible sur ordinateur, tellement là est une base idéale pour utiliser un ordinateur, les domaines ou les secteurs sur lesquels en rubrique/texte est orienté sont normalement "physiquement" marqués, de sorte que la classification ou même les fonctions de thésaurus puisse être employée particulièrement pour le transfert lexical (la désambiguisation), les besoins techniques de vocabulaire d'être très précis, de sorte que l' ordinateur aide en étant compatible.

Indexation automatique

L'indexation (le complet) du texte pourrait être un bon effet secondaire d'employer MT et CHAT. Pour le transfert lexical, on doit dériver des formes de mot aux formes de base; des composés et des mots complexes doivent être identifiés en tant que tel, exprimez les informations de classe, même des relations entre les termes sont employées pour la désambiguisation. Ainsi il peut être considéré de fournir de tels résultats de production ou d'intermédiaire des systèmes MT aux fins de l'archivage de document et la recherche documentaire.

Texte et traitement de textes

Il n'y a aucun doute que le traitement de textes joue un rôle important dans chaque environnement de traduction. Même les traducteurs (humains) freelance de plus en plus emploieront un système de traitement de textes (sur le PC), et là sont une petite étape pour intégrer, de quelque façon que ce soit, des glossaires ou des listes (privés) de mot accessibles par l'intermédiaire des soi-disant "fenêtres" sur l'écran — au lieu d'utiliser des boites d'carte-indice. Il est bien évident que d'autres fonctions, g de E.. les contrôleurs d'orthographe, de grammaire et de style, seront de plus en plus intégrés dans un tel processus.

En conséquence, le "texte source" est lisible sur ordinateur, mais on doit adapter MT et le système de CHAT (aux différents) systèmes de traitement de textes (par exemple le Wang OIS est combiné avec les LOGOS et SYSTRAN, WordPerfect est combiné avec SYSTRAN). Si de tels outils sont disponibles, la post-édition des résultats MT peut être soutenue par le rédacteur spécial.

Un problème dans cet environnement est la combinaison de ces outils avec les équipements locaux de MT (voir, par exemple, des LOGOS) ou de la connexion avec un centre de service de traduction (voir, par exemple , le concept de l'application SYSTRAN dans la Commission européenne ou même l'utilisation du minitel-système en France obtenir des traductions automatiques (par SYSTRAN).

EXTRACT FROM UNEDITED MACHINE TRANSLATION OF PAPER 8.

In the scientific and technical literature, especially in the fields of high technology, the terminology is moving Research, the industrial sector and the commercial sector are confronted with a discovered news, new products, therefore with new names. The lack of terminological standardization and the use of industrial jargons make comprehension even more difficult not only when it is foreign languages, but also for the users practising the language of the document.

The information science, created to the specialists'profit, tends to cover exhaustively the scientific and technical literature available in the world. The production of computerized bases and data banks became practical current and in fact the only solution of information vis-à-vis the increasing document mass.

Two are distinguished kind of bases and data banks: those which are of universal utility (CHEMABS, INSPEC, METADEX, NTIS, etc) and those which are worked out for use plus restraint, for specific industry, some undertakings or even only one institution. But whatever they are, their user producers and run up against the language and terminology problems.

Standardization terminological itself, as well as the production of the data bases, closely dependent on the communication between the document producers , to the availability of these documents, to the description of those and to the organization of coherent information.

The availability of the documents is problematic for well-known reasons:

— industrial concurence, safety, reason of state;
— difficulty of producers' tracking and holders of documents and of contact with those.

The compilation of a data base constitutes a first step towards a better communication and towards terminological standardization: indeed, the producer of the data base contacts the authors of the documents, gathers these documents or their description, indexes them in a uniform way, creating thus a more or less homogeneous terminology. In return, the users of the data bases familiarize themselves with only one terminology and tend to use it.

The universal use of the data bases by users of various countries and of languages various accustoms them all to only one language and only one terminology (in majority English) and already opens the bréche in the wall of linguistic incomprehension.

It is advisable here to stress that the linguistic problems are different from a country to another: there are unilingual and multilingual countries .

Languages also can be divided into two categories: easy said languages: Germanic Latin languages and, and languages that one describes as difficult to reach: languages Slavic, Chinese, Japanese, Arab, etc); of these last only will be retained those which convey currently a significant scientific production, such as for example Japanese and Russian. When one differentiates the problems by country one points out that the unilingual countries such as France, Great Britain, the USA have problems with all languages; but if they have difficulties of comprehension, on the other hand they for a long time set up structures and created organisms facilitating the translation for languages of both groups. The figures are significant: 29% of all the translations carried out concern the pairs of French-English and German-English languages.

The multilingual countries such as Switzerland, Canada, Belgium have much less problems of comprehension of the original texts and have consequently a network of scientific translation weaker, centred especially on difficult to reach languages.

To reduce the linguistic barrier do not mean to translate systematically the texts in the language of the user of the moment. When one speaks about the translation and structure network facilitating the translation, not very many are those which carry out of it the width and diversity. It is however of primary importance that the information users and the specialists in information know them.

There are initially the integral translations of documents (cover-to-cover translations) that one publishes especially for the difficult to reach language documents (Russian, Japanese). There is also reference works which point out them.

There is also a considerable number of texts which are already translated and can be re-used at a later date several times. For that one should detect them, to collect them and to compile them in the form of reference or data bank works to point out them to the potential users. The translated document collection comes itself up against the already evoked difficulties of competition, of safety.

It is in these particular cases that have to intervene of the more restricted groupings guaranteeing control of the users, defence of the interests of each one inside the grouping as well as professional secrecy for the translation suppliers and their users. It is only in this context, by respecting the evoked factors that it is possible to gather the translations already carried out.

It goes without saying that the data base is the most valid solution to put these translations at the disposal of the interested parties. For example it is advisable to quote "World Translation index", data base which proposes the translations gathered in Europe and in the USA by the translation national centres (USA) and the international translation centres (Europe). WTI is also diffused in reference traditional work form.

Certain groupings point out for their members the translations carried out within the group, for example the metallurgical research centre, Belgium.

As a last resort to abolish the linguistic barrier there is obviously the translation. One distinguishes:
- traditional translation.
- translation carried out by computer;
- computer-assisted translation.

The method of traditional translation called upon the translators and presents certain difficulties, especially for the very specialized technical texts: the translator encounters problems of technical comprehension and terminology.

The translation by computer made major progress, but its effectiveness and the quality of the translation is still debatable. Multiple studies spoke about it, evaluating its advantages and its defects. It is considered however that these translations are valid at least for the scientists' information and provided that the vocabulary input into the computer is limited to a properly determined field. It is useful to stress the positive result obtained for the indexdation of the documents and the preparation of the summaries to introduce into the data bases. The standard or repetitive texts are also rather well translated . The principal advantage of the translation by computer is indisputably the possibility of standardization of the terminology. The phenomenon of standardization caused by the data bases for language appears here for several languages. Unfortunately the essential reproaches that is generally formulated are considerable: price too high, considerable waste of time in revision of the texts, inaptitude of the computer to take into consideration the nuances of the text.

Computer-assisted translation, as for more flexible it, combine the creative the translator's work and the possibilities of data processing. It is for that that it presents an advantageous, valid, realizable solution without excessive expenses. Its advantages are obvious for the user of the texts and for the translator, the output and the quality of the translation are improved.

It is noted that the possibilities of reduction of the linguistic barrier are numerous; it will be therefore useful to know the institutions which implement them.

There are the translation and diffusion organisms above all of the bases and data banks:

- base producers and data banks;
- information centres;
- description centres of the translations;
- terminological centres.

Organisms contact the document producers, compile the literature, put information at the disposal of the user, gather the existing translations, produce and exploit computerized terminological dictionaries. They are useful to the user since they answer in various ways their requirements in information and abolish the linguistic barrier. They are also the auxilliaires of the translator since they are implied in the production and the exploitation of the terminological sources and of the sources of information on the existing translations.

Information user confronted with the linguistic problems and the translator who has to help it has both the major interest to know them and to resort to their services.

Nevertheless, when it is a question of abolishing the barrier linguistic he is well the translator who is personnage-clé it. These problems deserve consequently a detailed study.

More than ever the translator finds himself confronted with very specialized scientific texts and with descriptions of high technicality. He is compelled to cope with the problems:
- comprehension of the texts;
- terminological research for advanced techniques;
- research of the sources of information;
- an adaptation to the modern work and information retrieval methods.

The comprehension of the texts is made difficult by the complexity of the subjects which for 55 % raise field of precise sciences and of the high technology; the very pushed specialization of the disciplines implied in each subject insists on a certain scientific training or at least a close contact with the specialist.

The research of the equivalent terms in the source language and in the target language is made extrèmement difficult because of mouvance of the terminology of the lack of standardization in the definitions of new products and methods, industrial jargons: the translator devotes on average 17 % of his time in the search for information and 58 % only to the translation as such.

The dictionaries — traditional the translator's tools — present a considerable shift in relation to the proposed texts, especially in the fields mentioned and in advanced disciplines.

The petrol even of the dictionary and the edition methods does not allow any sufficiently rapid update; it is generally considered that they are 4 years of delay on the technological developments.

BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION

**Technical Information Panel Specialists' Meeting,
Washington, DC, USA, 23–24 September, 1987**

LIST OF PARTICIPANTS

Mr A.AINES	Consultant, 5623 Bellington Ave., Springfield, VA 22151, USA
Mr C.L.ALLEN	Liaison Officer, Canadian Defence Liaison Staff, 2450 Massachusetts Ave. N.W., Washington DC 20008, USA
Ms N.J.ALLIN	Chief, STINFO Office, Air Force Human Resources Laboratory, STINFO Office (TSR), Brooks AFB, TX 78235-5601, USA
Mr D.A.APPLER	Asst. STINFO, Secretary of Defense, The Pentagon, Rm 3E114, Washington, DC 20301-3082, USA
Miss E.AXIOTOPOULOU*	Hellenic Air Force General Staff, C Branch C1/3, Holargos/Athens, Greece
Dr T.BELLARDO	Asst. Professor, Catholic University of America, 629 Michigan Ave. N.E., Washington DC 20064, USA
Mr R.BERNHARDT*	GMD-F4, Braunshardter Weg 11, D-61100 Darmstadt, FRG
Ms R.BIXBY	Technical Info. Specialist, Defense Tech. Info. Center, Cameron Station DTIC-EB, Alexandria, VA 22304-6145, USA
Mr W.R.BLADOS*	USAF Scientific & Technical Information Program Manager, Office of the Secretary of the Air Force, SAF/AQXT, Room 4C258, The Pentagon, Washington DC 20330-1000, USA
Prof. D. BODDY†	Department of Management Studies, University of Glasgow, 57 South Park Avenue, Glasgow, UK
Mr J.S.BORTMAN	Technical Transfer Coordinator, Naval Air Development Center, Warminster, PA 18974, USA
Ms J.BORTNICK†	Asst.Chief, Science Policy Research Division, Congressional Research Service, Library of Congress, Washington, DC 20540, USA
Mr C.P.BOURNE	Director, General Information Div., Dialog Information Services, Inc., 3460 Hillview Avenue, Palo Alto, CA 94304, USA
Mr M.BRANDRETH*	Chief, Policy, Planning & Systems, Canada Institute for Scientific and Technical Information, National Research Council of Canada, Ottawa, Ontario K1A 0S2, Canada
Lt. Col. H.BRAUN*	Dokumentationszentrum Bw, Dezernat DOK-Info, Friedrich-Ebert Allee 34, 5300 Bonn 1, FRG
Ir. G.M.BREAS*	Director, Scientific & Technical Documentation Centre for the Armed Forces (TDCK), Postbus 90701, 2509 LS The Hague, The Netherlands
Mr W.S.BURGMAN	Director, Air Weather Service Technical Library, Scott AFB, IL 62225-5458, USA
Dr H.D.BURTON	Technical Information Specialist, Lawrence Livermore National Lab., P.O. Box 808, L-542, Livermore, CA 94550, USA
Dr J.F.CAPONIO	Director, National Technical Info. Service, 5285 Port Royal Road, Springfield, VA 22161, USA
Ms J.CATHCART*	Acting Director, Directorate of Scientific Inf. Services, Research and Development Branch, National Defence Headquarters, 101 Colonel By Drive, Ottawa, Ontario K1A 0K2, Canada
Ing. Gen. F.CHEVALIER*	Directeur, C.E.D.O.C.A.R., 26 Boulevard Victor, 75996 Paris Armées, France
Mr J.P.CHILLAG	Special Acquisitions, British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ UK
Mrs I.CLEMENS†	Royal Library CBT, Boulevard de l'Empereur, 1004 Brussels, Belgium
Mr E.D.COLLINS	Chief, Technical Info. Branch, Defense Logistics Agency, Attn: DLA-SCT, Cameron Station, Alexandria, VA 22304-6100, USA

†Author
*Panel Member

Dr Ing. A.M.RAMALHO CORREIA*	L.N.E.T.I., Director, C.I.T.I., Azinhaga dos Lameiros, 1699 Lisboa, Codex, Portugal
Ms G.A.COTTER†	Chief, Info. Systems Division, Defense Technical Info. Center, Cameron Station, Attn: DTIC-EB, Alexandria, VA 22304-6145, USA
Dr K.CROUCH	Chief, Librarian, Royal Military College of Canada, Kingston, Ontario K7K 5L0, Canada
Lt. Col. A.CUFFEZ*	Adjoint Informatique Logistique, Etat-Major Force Aérienne (VSL/I), Quartier Reine Elisabeth, Rue D'Evère, 1140 Brussels, Belgium
Mr D.C.CUNNINGHAM	Nuclear Engineer/Defense Programs, US Department of Energy-OSTI, P.O. Box 62, Oak Ridge, TN 37831, USA
Mrs C.DEGNAN	AFAL/STINFO Officer, Edwards AFB, CA 93523-5000, USA
Ms D.DELEUZE*	Information Specialist, Construcciones Aeronauticas S.A. (CASA), R&D Technology and Materials Dept., Madrid, Getafe, Spain
Mr D.R.DRESIA	Manager Director/PUBS, American Society of Civil Engineers, 345 East 47th St., New York, NY 10017, USA
Dr N.S.DUMAS	Director, Policy & Plans, Attn: DTIC, Cameron Station, Alexandria, VA 22304-6145, USA
Ms C.DURANCE†	National Library of Canada, Wellington Street, Ottawa, Ontario K1A ON4, Canada
Mrs T.J.EGAN	Information Specialist, Department of Energy, Office of Military Application, Washington, DC 20545, USA
Mr J.J.ELSBREE	Office of Publications, NTIS, 5285 Port Royal Road, Springfield, VA 22161, USA
Col. Dr R.FOESEL	Head, DOKZENTBw, Friedrich-Ebert-Allee 34, D-5300 Bonn 1, FRG
Mr J.L.FORMAN	Marketing Manager, Derwent, Inc., 6845 Elme Street, McLean, VA 22101 USA
Ms B.FOX	Special Assistant, Defense Technical Info. Center, Cameron Station, Alexandria, VA 22304-6145, USA
Ms P.M.GAYNOR	Director, Office of Systems Modernization, Defense Technical Info. Center, Attn: DTIC-R, Alexandria, VA 22304-6145, USA
Mr J.GEORGET	Engineer, Aérospatiale, B.P. 96, 78133 Les Mureaux Cedex, France
Mr J.E.GIGNAC	Deputy General Manager, Mission Support, NASA Scientific & Tech. Inf. Facility, P.O. Box 8756, BWI Airport, MD 21240, USA
Mr D.GILLILAND	Computer Scientist, Defense Technical Info. Center, Attn: DTIC-EB, Cameron Station, Alexandria, VA 22304-6145, USA
Mrs R.A.GJERSVIK*	Director of the Library, The Technical University, Library of Norway, Hogskoleringen 1, N-7034 Trondheim — NTH, Norway
Captain A.F.VAN DE GRAMPEL	NATO/IMS, International Military Staff, Armaments & Standardisation Division, B-1000 Brussels, Belgium
Mr D.U.GREEVY	Director, Federal Systems, Telebase Systems, 5403 Midship Court, Burke, VA 22015, USA
Mr V.E.HAMPEL	Program Leader for Technology & Inf. Systems, Lawrence Livermore National Lab., L-542 P.O. Box 808, Livermore, CA 94550, USA
Ms M.HANNA	Program Analyst, Defense Technical Info. Center, Cameron Station, Alexandria, VA 22304-6145, USA
Mr A.J.HANSBROUGH	Chief, RIAD (NASA LaRC), M/S 180A, NASA Langley Research Center, Hampton, VA 23665-5225, USA
Mr J.J.HARFORD†	Executive Director, American Institute of Aeronautics and Astronautics, 370 L'Enfant Promenade, SW, Washington DC 20024, USA
Mr H.HEYCK	Manager, Library & Information Centre, Transpatt Canada, Floor 2, Tower 'C', Place de Ville, Ottawa, Ontario K1A ON5, Canada
Mrs G.HODGE	Head, R&D, BIOSIS, 2100 Arch St., Philadelphia, PA 19103, USA

†Author
*Panel Member

Ms C.HOLLANDER	Threat Analyst, United Technologies Corp., 2000 Corporate Ridge, Suite 850, McLean, VA 22102, USA
Mr D.HOLZ	Technical Info. Officer, US Department of Energy, 1333 Broadway, Oakland, CA 94612, USA
Mr R.HORNE	Head of Acquisitions, British Library, Boston Spa, Wetherby, West Yorkshire, LS23 7BQ, UK
Dr J.M.HUGHES II	Head, Technical Library, Naval Surface Weapons Center, Code E23, Dahlgren, VA 22448, USA
Mr P.H.IMHOF	Technical Info. Officer, Naval Research Lab, Washington DC 20375, USA
Mrs M.J.JOHNSON	Technical Librarian, FluiDyne Engineering Corp., 5900 Olson Memorial Highway, Minneapolis, MN 55422, USA
Lt. Col. C.M.JORGE*	CLAFA-DMA, Base de Alfragide, Av. Leite de Vasconcelos, 2700 Amadora, Portugal
Mr A.D.KAHN	Project Manager, Defense Technical Info. Center, Cameron Station, (DTIC-EB), Alexandria, VA 22304-6145, USA
Col. D.KAYA*	Ministry of National Defence (MSB), Dept. of R&D (ARGE), Ankara, Turkey
Mr M.J.KERPER	Chief, Technical Info. Division, AFSOR/XOT, Building 410, Bolling AFB, DC 20332-6448, USA
Mr R.KHAN	Executive Secretary, CENDI, NTIS, 5285 Port Royal Road, Room 200F, Springfield, VA 22161, USA
Dr D.W.KING	President, King Research Inc., 6000 Executive Blvd., Rockville, MD 20852, USA
Mr P.KLINEFELTER	Program Manager for Information, Analysis Centers, DTIC-DF, 325 Pitt St., Alexandria, VA 22314, USA
Mr R.LARUE†*	Coordinateur de l'Information Scientifique et Technique du CEA, Centre d'Etude Nucléaires de Saclay, 91191 Gif sur Yvette, France
Mr O.LAVROFF*	Société Aérospatiale, Chef du Dept. Information et Documentation, B.P. No.76, 92152 Suresnes, Cedex, France
Ms B.LAWRENCE†*	Administrator, Technical Information Service, American Institute of Aeronautics and Astronautics, 555 West 57th Street, 12th Floor, New York, NY 10019, USA
Ms P.LISTER	Manager, Air Force Information for Industry Office, 5001 Eisenhower Avenue, Alexandria, VA 22304
Mr J.A.LOWE*	Chief Technical Librarian, British Aerospace plc, Military Aircraft Division, Warton Aerodrome, Preston, Lancs PR4 1AX, UK
Dr R.B.MAHONEY	Deputy Manager, DASIAC/KSC, 250 Huntington Ave, Suite 500, Alexandria, VA 22303, USA
Mrs D.H.MARTIN	Technical Information Program Manager, US Department of Energy/CTID, P.O. Box 5400, Albuquerque NM 87115, USA
Ms E.V.MCCAULEY	Director, DTIC-E, Defense Technical Information Center, Cameron Station, Attn: DTIC-E, Alexandria, VA 22304-6145, USA
Miss S.G.MCINTYRE	A/DSIS-2, CRAD/DSIS, 10 Colonel By Drive, Ottawa, Ontario K1A 0K2, Canada
Mr H.W.MILES	Vice President, Mission Support, Planning Research Corporation, 1505 Planning Research Drive McLean, VA 22102, USA
Mrs M.D.MILLER	Chief Librarian, Main Library, Bldg, F6.2, Atomic Weapons Research Establishment, Aldermaston, Reading RG7 4PR, UK
Ms R.MILLER	Engineering Information, 345 East 57th St., New York, NY 10017, USA
Mr J.B.MILSTEAD	Director, Telecommunications and ADP Systems, DTIC-2, Cameron Station, Alexandria, VA 22304-6145, USA
Mr K.MOLHOLM*	Administrator, Defense Technical Information Center, Cameron Station, Alexandria, VA 22304, USA

†Author
*Panel Member

Mr O.OEGEMA	Mgr. Info. Center, Dow Chemical (Nederland) B.V., P.O.Box 48 Terneuzen, 4530 AA, The Netherlands
Mrs D.PATRINOU*	Technical Scientific Library, Hellenic Aerospace Industry, P.O.Box 23, GR 32009 Schimatari, Viotias, Greece
Mr G.E.C.PRATT	Le Chataignier, Pyrford Road, Woking, Surrey, UK
Eng. R.M.S.PEREIRA	Head, Doc. Center, Av. Luis de Camões, Lote, 1 Bloco C-9, 8° D, 2800 Miratejo, Portugal
Mr H.E.PRYOR	Logistics Management, Inc, 6400 Goldsboro Road, Bethesda, MD 20817, USA
Ms M.A.PUTNAM	Chief, Technical Reports Branch, AFWL/SUR, Kirtland AFB, NM 87117-6008, USA
Mr P.A.ROBEY, Jr	Dept. Administrator, Attn: DTIC, Cameron Station, Room 5D121, Alexandria, VA 22304-6145, USA
Mr R.P.RYAN	Director, Office of User Services & Marketing, Defense Technical Information Center, Alexandria, VA 22304-6145, USA
Mr H.SAUTER	Retired Administrator, Defense Technical Information Center, SR4 Box 25k, Gunsmoke Trail, Lusby, MD 20657, USA
Ms S.A.SAVAGE†	Supervisor, ARCOvision Services, Atlantic Richfield Company, 1500 Market Street, Philadelphia PA 19102, USA
Mr R.SEARLE*	Chief Librarian, Royal Aircraft Establishment, Farnborough, Hants GU14 6TD, UK
Mrs R.SMITH	Director of Customer Services, Forbes Place, Springfield, VA 22161, USA
Mr F.SOBIESZCZYK	Chief, Research Program Officer, ODUSD, Pentagon, Room 3E114, Washington DC 20301, USA
Mr W.SPANOS	Project Manager, Defense Technical Information Center, Cameron Station, Alexander, VA 22304-6145
Mr E.J.SPICER	Parliamentary Librarian, Library of Parliament, Parliament Bldgs, Ottawa, Ontario K1A 0A9, Canada
Ir. A.S.T.TAN*	Information Specialist, National Aerospace Laboratory (NLR), P.O.Box 90502, 1006 BM Amsterdam, The Netherlands
Mr P.A.THIBIDEAU	Manager, International Scientific & Technical Information Activities, NASA Headquarters (NTT-4), Washington DC 20546, USA
Mr W.M.THOMPSON	Director, Database Services, Defense Technical Information Center, Cameron Station, Alexandria, VA 22304-6145, USA
Mr G.TITTLBACH*	Fachinformationszentrum, Karlsruhe, D-7514 Eggenstein-Leopoldshafen 2, FRG
Miss B.UNRUH	Executive Director, National Federation of Abstracting, Information Services 1429 Walnut Street, Philadelphia, PA 19102, USA
Mrs E.URUNDUL*	TUBITAK-TURDOK, Tunus Caddessi 33, Kavaklidere, Ankara, Turkey
Mr G.VANAUTRYVE*	Scientific Advisor, Centre National de Documentation Scientifique & Technique, Bibliothèque Royale Albert 1er, Bld. de l'Empereur 4, B-1000 Brussels, Belgium
Miss C.WALKER*	Head, Information Services Branch, SHAPE Technical Centre, P.O.Box 174, 2501 CD The Hague, The Netherlands
Dr T.E.WALSH	Director of Sponsored Research, 219 Grinter Hall, University of Florida, Gainesville, FL 32611, USA
Mrs L.WARGO	STINFO Officer, AFATL/DOI/STINFO, Eglin AFB, Fl 32542-5434, USA
Mr M.I.WEINBERGER	Senior Vice President, Telebase Systems, 763 W. Lancaster Avenue, Bryn Mawr, PA 19010, USA
Mr V.A.WENTE*	Director, Scientific & Technical, Information Division, NASA Headquarters (Code NTT), Washington, DC 20546, USA
Mr M.R.C.WILKINSON	Head, Defence Research Information Centre, Kentigern House, 65 Brown Street, Glasgow G2 8EX, UK
Ms B.WILLIAMS	STINFO Officer, HSD/SORT, Brooks AFB, TX 78235, USA

†Author
*Panel Member

Mr J.WILSON	Head, Database Product & Services, NASA STINFO Office, Code NTT, Washington, DC 20546, USA
Mr D.L.WOODS	Director, Navy Scientific and Technical Information Office, Chief of Naval Research, P.O.Box 15114, Arlington, VA 22215, USA
Mr A.YANEZ	C.E.D.O.C.A.R., 26 Boulevard Victor, 75996 Paris Armées, France
Ms S.E.YOUNG	Assistant Director for Technical Information, Defense Nuclear Agency, 6801 Telegraph Rd, Alexandria, VA 223100, USA
Prof. Dr H.ZIMMERMANN†	Universität des Saarlandes, Fachrichtung 5.5, Dienstgebaude 4, D-6600 Saarbrucken 11, FRG

†Author

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-430	ISBN 92-835-0449-6	UNCLASSIFIED

5. Originator	Advisory Group for Aerospace Research and Development North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly sur Seine, France
6. Title	BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION
7. Presented at	the Technical Information Panel Specialists' Meeting held in Washington, DC, USA on 23—24 September 1987.

8. Author(s)/Editor(s)	9. Date
Various	March 1988
10. Author's/Editor's Address	**11. Pages**
Various	116

12. Distribution Statement	This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Information	Meetings
Barriers	Aerospace industry
Communications management	Military communication
Internal relations	

14. Abstract

These Conference Proceedings comprise the papers presented at the Specialists' Meeting held by the Technical Information Panel in Washington, DC, USA, 23—24 September 1987. Topics discussed include an overview of the barriers to the international transfer of information, particularly in aerospace and defence; linguistic, political and economic barriers; the use of technology to overcome these and other barriers, e.g. machine translation, links between disparate databases and tele-conferencing; the importance of standards; and the need to change organisations when introducing or changing technology to prevent internal barriers reducing the acceptance and use of the new technology.

AGARD Conference Proceedings No.430 Advisory Group for Aerospace Research and Development, NATO BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION Published March 1988 116 pages These Conference Proceedings comprise the papers presented at the Specialists' Meeting held by the Technical Information Panel in Washington, DC, USA, 23–24 September 1987. Topics discussed include an overview of the barriers to the international transfer of information, particularly in aerospace and defence; linguistic, political and economic barriers; the use of technology to overcome these and other barriers, e.g. machine translation, P.T.O.	AGARD-CP-430 Information Barriers Commmunications management Internal relations Meetings Aerospace industry Military communication	AGARD Conference Proceedings No.430 Advisory Group for Aerospace Research and Development, NATO BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION Published March 1988 116 pages These Conference Proceedings comprise the papers presented at the Specialists' Meeting held by the Technical Information Panel in Washington, DC, USA, 23–24 September 1987. Topics discussed include an overview of the barriers to the international transfer of information, particularly in aerospace and defence; linguistic, political and economic barriers; the use of technology to overcome these and other barriers, e.g. machine translation, P.T.O.	AGARD-CP-430 Information Barriers Communications management Internal relations Meetings Aerospace industry Military communication
AGARD Conference Proceedings No.430 Advisory Group for Aerospace Research and Development, NATO BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION Published March 1988 116 pages These Conference Proceedings comprise the papers presented at the Specialists' Meeting held by the Technical Information Panel in Washington, DC, USA, 23–24 September 1987. Topics discussed include an overview of the barriers to the international transfer of information, particularly in aerospace and defence; linguistic, political and economic barriers; the use of technology to overcome these and other barriers, e.g. machine translation, P.T.O.	AGARD-CP-430 Information Barriers Communications management Internal relations Meetings Aerospace industry Military communication	AGARD Conference Proceedings No.430 Advisory Group for Aerospace Research and Development, NATO BARRIERS TO INFORMATION TRANSFER AND APPROACHES TOWARD THEIR REDUCTION Published March 1988 116 pages These Conference Proceedings comprise the papers presented at the Specialists' Meeting held by the Technical Information Panel in Washington, DC, USA, 23–24 September 1987. Topics discussed include an overview of the barriers to the international transfer of information, particularly in aerospace and defence; linguistic, political and economic barriers; the use of technology to overcome these and other barriers, e.g. machine translation, P.T.O.	AGARD-CP-430 Information Barriers Communications management Internal relations Meetings Aerospace industry Military communication

links between disparate databases and tele-conferencing; the importance of standards; and the need to change organisations when introducing or changing technology to prevent internal barriers reducing the acceptance and use of the new technology. ISBN 92-835-0449-6	links between disparate databases and tele-conferencing; the importance of standards; and the need to change organisations when introducing or changing technology to prevent internal barriers reducing the acceptance and use of the new technology. ISBN 92-835-0449-6
links between disparate databases and tele-conferencing; the importance of standards; and the need to change organisations when introducing or changing technology to prevent internal barriers reducing the acceptance and use of the new technology. ISBN 92-835-0449-6	links between disparate databases and tele-conferencing; the importance of standards; and the need to change organisations when introducing or changing technology to prevent internal barriers reducing the acceptance and use of the new technology. ISBN 92-835-0449-6

AGARD

NATO OTAN

7 rue Ancelle · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone (1)47.38.57.00 · Telex 610 176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres.Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Board
Ved Idraetsparken 4
2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon

GERMANY
Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kar sruhe
D-7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Aircraft Support Equipment Directorate
Department of Research and Development
Holargos, Athens, 15500

ICELAND
Director of Aviation
c/o Flugrad
Reyjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 AC Delft

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Portuguese National Coordinator to AGARD
Gabinete de Estudos e Programas
CLAFA
Base de Alfragide
Alfragide
2700 Amadora

TURKEY
Millî Savunma Bakanlığı (MSB)
ARGE Daire Başkanlığı (ARGE)
Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Research Center
M/S 180
Hampton, Virginia 23665

THE UNITED STATES NATIONAL *DISTRIBUTION CENTRE (NASA) DOES NOT HOLD* STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
England

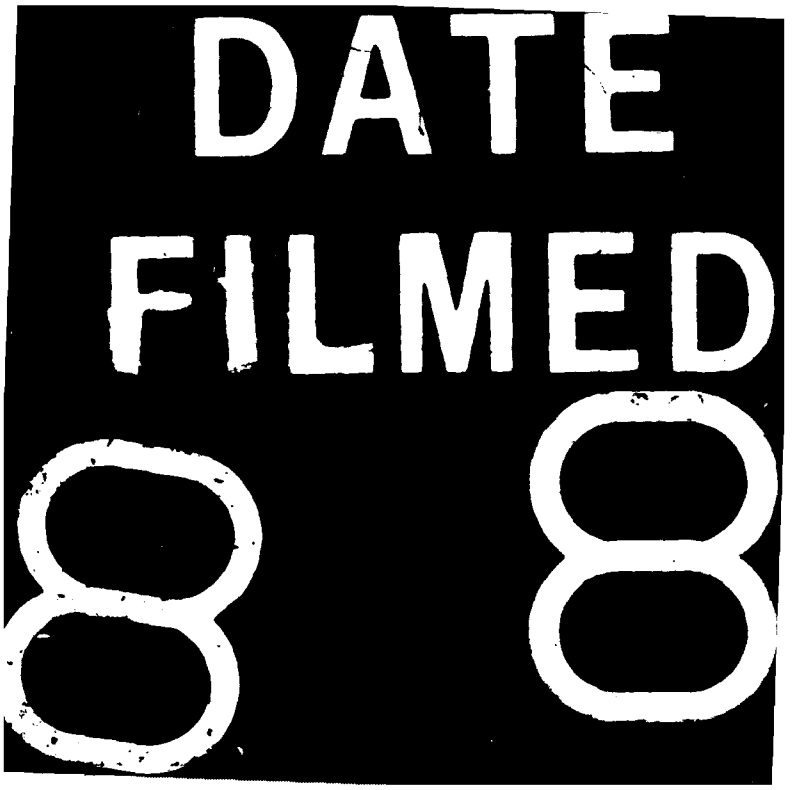
Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Branch
NASA Headquarters (NIT-40)
Washington D.C. 20546, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

ISBN 92-835-0449-6


DATE
FILMED
8 8